

Marko Myllymäki

**Kvantitatiivisen urapolkuanalyysin
kehittäminen ja arviointi:
suomalaisten
bioteknologiaturkijoiden uratyypit**

Teknillinen korkeakoulu
Tietotekniikan osasto

Helsinki University of Technology
Department of Computer Science and Engineering

TEKNILLINEN KORKEAKOULU
Helsinki ja ympäristötekniikan
osaston kirjasto

TEKNILLINEN KORKEAKOULU DIPLOMITYÖN TIIVISTELMÄ

Tekijä:	Marko Myllymäki	
Työn nimi:	Kvantitatiivisen urapolkuanalyysin kehittäminen ja arviointi: suomalaisten bioteknologiaturkijoiden uratyypit	
Päivämäärä:	8.11.2004	Sivuja: 71+21
Osasto:	Tietotekniikan osasto	Professuuri: Yhd-102
Työn valvoja:	Professori Janne Hukkinen	
Työn ohjaaja:	Tohtoritutkija Henrik Bruun	

Diplomityössä on toteutettu Matlab-ohjelmistolla kvantitatiivinen analyysikehikko urapolkujen tutkimiseen, jossa menetelminä on käytetty pääkomponenttianalyysia (PCA) ja verkostanalyysia. Työn pääpainoalue on ollut erityisesti menetelmien kehittämisessä urapolkuaineiston tutkimiseen, ja näin ollen hypoteeseja ei ole asetettu, eli analyysi on luonteeltaan kartoittavaa.

Analyysimenetelmien testausaineistona on käytetty suppeaa joukkoa suomalaisten bioteknologia-alan avaintoimijoiden urapolkutietoja, jotka on kerätty ansioluettelo-muodossa. Alkuperäinen suunnitelma tutkia bioteknologia-alan yritysjohtajien urapolkuja jouduttiin hylkäämään aineiston vaikean saatavuuden takia.

Pääkomponenttianalyysin avulla on tutkittu, kuinka kvantitatiivisesta uradatasta voidaan hahmottaa keskeisiä uratyyppejä. Moniulotteinen data on projisoitu menetelmäs-sä pienempiulotteiseen avaruuteen niin, että suurimman varianssin omaava informaatio on saatu esille.

Verkostonalyysissä on tarkasteltu henkilöiden urapolkujen ammatillisia kohtaamis-pintoja. Tällä analyysitavalla havaittiin olevan mahdollista löytää merkittäviä toimijoi-ta ja organisaatioita alalla.

Työn tulos on kvantitatiivinen analyysimalli, joka toimii lähtökohtana urapolkujen kar-toitukseen. Tulosten arvioinnissa havaittiin, että aineistokysymyksiin on kiinnitettävä erityistä huomiota, liittyen aineiston kattavuuteen ja tarkkuuteen.

Avainsanat: urapolut, bioteknologia-ala - Suomi, kvantitatiivinen analyysi, pääkompo-nenttianalyysi, verkostonalyysi

HELSINKI UNIVERSITY
OF TECHNOLOGY

ABSTRACT OF THE
MASTER'S THESIS

Author:	Marko Myllymäki	
Name of the thesis:	Development and Assessment of Quantitative Career Path Analysis: Career Types of the Finnish Biotechnology Researchers	
Date:	November 8, 2004	Number of pages: 71+21
Department:	Computer Science and Engineering	Professorship: Yhd-102
Supervisor:	Professor Janne Hukkinen	
Instructor:	Researcher Henrik Bruun, PhD	
<p>In the thesis, quantitative analysis framework for studying career paths has been developed with Matlab computing software. Methods for the analysis were the principal component analysis (PCA) and the network analysis. The main focus of the thesis has been the development of methods for analyzing career path data. Therefore no hypotheses has been stated and the study is exploratory in nature.</p> <p>Narrow set of Finnish biotechnology actors' resumés has been used to test the quantitative methods. The original idea of studying career paths of CEO's of Finnish biotechnology enterprises was given up because of difficulties getting the data.</p> <p>The principal component analysis has been used to perceive general patterns of careers from the data. High dimensional data has been projected to lower dimensions in order to get maximum variance information.</p> <p>The network analysis has been used to examine connections in actors' career paths. With this analysis it turned out to be possible to find significant actors or organizations in the field.</p> <p>The result of the thesis is quantitative analysis framework, which can be the basis for studying career paths. It was found out in the evaluation part that accuracy and scope of the data must be carefully concerned.</p>		
Keywords: career paths, biotechnology sector - Finland, quantitative analysis, principal component analysis, network analysis		

Esipuhe

Tämä diplomityö on tehty Teknillisen korkeakoulun Ympäristönsuojelulaboratoriossa. Työ liittyy kahteen laboratorion bioteknologia-alaa käsittelevään tutkimushankkeeseen, joista toisessa on tutkittu osaamisverkostojen hallintaa innovaatioprosesseissa (Managing Transepistemic Innovations) ja toisessa on tutkittu suomalaisen bioteknologian sosio-kulttuurisia ulottuvuuksia (Socio-cultural Dimensions of Technological Change: The Case of Finnish Biotechnology).

Kiitän Ympäristönsuojelulaboratorion professori Janne Hukkista sekä tohtoritutkija Henrik Bruunia mielenkiintoisesta aiheesta ja mahdollisuudesta työskennellä tutkimusapulaisena tämän projektin puitteissa. Kiitän koko laboratorion tutkijakuntaa kannustavasta suhtautumisesta ja saamistani kommentteista.

Kiitokseni kuuluu kaikille tutkimukseen osallistuneille, jotka ystävällisesti luovuttivat henkilökohtaisia uratietojaan tutkimuskäyttöön. Kiitän myöskin kaikkia asiantuntijoita, joilta olen työn eri vaiheissa saanut kommentteja.

Lisäksi haluan kiittää kotiväkeäni saamastani kannustuksesta. Lämpimin kiitokseni kuuluu rakkaalle vaimolleni Kaisalle, joka antoi arvokkaita huomautuksia työn rakenteeseen ja kieliasuun liittyen, sekä tuki työn kaikissa vaiheissa.

Helsingissä, 8 marraskuuta, 2004

Marko Myllymäki

Sisältö

Kuvat	7
Taulukot	9
1 Johdanto	10
1.1 Biotekniikan ja bioteknologian määritelmistä	11
1.2 Bioteknologia Suomessa	13
1.3 Uratutkimuksista	15
1.4 Tutkimusraportin rakenne	16
2 Peruskäsitteet ja tutkimustehtävä	17
2.1 Käsitteet tutkimuksessa	17
2.2 Tutkimuksen merkitys ja tavoitteet	21
2.3 Tutkimustehtävä	21
3 Analyysin toteutus	23
3.1 Aineisto	24
3.2 Validiteetti ja reliabiliteetti	32
3.2.1 Yleistettävyyys	33
3.2.2 Otanta	33

3.2.3	Tämän tutkimuksen erityiskysymyksiä	35
3.3	Aineiston esikäsittely	36
3.3.1	Tarvittavien muuttujien valinta	36
3.3.2	Aineiston kvantifiointi muuttujien arvoiksi	41
3.4	Analyysiohjelmisto	43
3.5	Perustulokset kohdejoukosta	45
3.6	Pääkomponenttianalyysi uratyyppien hahmottamisessa	47
3.6.1	Menetelmän teoreettinen tarkastelu	47
3.6.2	Tarvittavien muuttujien konstruointi	50
3.6.3	PCA-analyysin toteutus	51
3.7	Verkostoanalyysi urapoluista	56
3.7.1	Lähtökohdat	56
3.7.2	Verkostoanalyysin käsitteitä	56
3.7.3	Analyysin toteutus	57
4	Tutkimuksen arviointi	64
4.1	Tavoitteiden arviointi	64
4.2	Jatkotutkimusmahdollisuudet	66
5	Yhteenveto	67
	Lähteet	71
A	Matlab- analyysiprosessien lähdekoodi	72
B	Kuvaus luokittelutyökalusta	88

Kuvat

2.1	Rinnakkaiset ura-askeleet	19
2.2	Ura-askelten ketju	19
2.3	Organisaatioliikkuvuus	20
3.1	Analyysin toteutus	24
3.2	Visualisointi tietorakenteista	40
3.3	Yksittäisen ura-askeleen tietorakenne A_i	41
3.4	Kohdejoukon ikäjakauma	45
3.5	Kohdejoukon koulutusjakauma	46
3.6	Kohdejoukon viimeisen työpaikan jakauma	46
3.7	Kaksiulotteinen esimerkki	49
3.8	PCA-analyysin tulokset	51
3.9	Dendrogrammi $N = 66$	53
3.10	PCA-analyysin tulokset, klusterit	54
3.11	Nelisolmuinen graafi - yhteydet	58
3.12	Henkilöt graafin solmuina	60
3.13	Henkilön X yhtäaikaiset organisaatioyhteydet	61
3.14	Histogrammi henkilöyhteyksistä	62
B.1	Luokittelusovellus - perustietojen lisäys	89

B.2	Luokittelusovellus - urapolun syöttäminen	90
B.3	Luokittelusovellus - luokat	91
B.4	Luokittelusovellus - lista henkilöistä	92

Taulukot

3.1	Fiktiivinen esimerkki ansioluettelosta, olennaiset tiedot	26
3.2	Henkilön perustiedot P_i	38
3.3	Yksittäisen ura-askeleen tiedot A_j	39
3.4	Analyysin Matlab-tiedostot	44
3.5	Havaintomatriisi	50
3.6	Redusoitu havaintomatriisi	51
3.7	PCA-analyysin tulokset	55

Luku 1

Johdanto

Diplomityössä tutkitaan suomalaisen bioteknologia-alan toimijoiden urapolkujen samankaltaisuuksia ja urapoluista muodostuvia verkostoja kvantitatiivisesti. Työn tavoitteena on kehittää urapolkujen tutkimusmenetelmä, jota voidaan jatkossa soveltaa laajempien aineistojen kohdalla hypoteesien testaamisessa. Kvantitatiivisten menetelmien käyttö urapolkujen analysoinnissa mahdollistaa aineiston tilastollisen käsittelyn ja yleistysten tekemisen sekä näistä tehtyjen luotettavuusarvioiden laskemisen. Yleistysten tekemiseen tarvitaan riittävän hyvä otos. Diplomityössä kehitetään tietokonesovellus, jolla voidaan toteuttaa suurten urapolkuaineistojen analysointi. Tämä työkalu toteutetaan numeeriseen laskentaan käytettävällä Matlab -ohjelmistolla.

Aineiston analysoinnissa selvitetään säännönmukaisuuksia, joiden voidaan katsoa olevan kiinnostavia mahdollisina jatkotutkimushypoteeseina ja yleisesti merkillepantavia suomalaisen bioteknologia-alan toimijoiden ja urapolkujen kartoituksessa. Tällaisia ovat merkittävien organisaatioiden kartoittaminen, aineistosta ilmenevät urapolkujen tyypit sekä ammatilliset kohtaamispiinnat eri toimijoiden välillä. Tässä tutkimuksessa käsiteltävä aineisto ei kuitenkaan ole edustava, ja näin ollen saatujen tulosten pohjalta ei tehdä yleistyksiä. Teoriatasolla tuodaan kuitenkin tähän liittyvät näkökohdat esille.

Diplomityöhön kuuluu urapolkuanalyysin kehittäminen ja sen arviointi. Lisäksi esitetään menetelmän tulokset empiirisen aineiston pohjalta. Arviointi tehdään aineis-

tolla, joka sisältää joukon suomalaisen bioteknologia-alan merkittäviä toimijoita.

Seuraavassa käydään läpi, miten bioteknologiaa on yleisesti käsitteenä pyritty määrittelemään, ja mikä on tämän tutkimuksen raja bioteknologialle. Lisäksi esitetään tiivis katsaus suomalaiseen bioteknologiakenttään tutkimuksen empiirisen kohteen selvittämiseksi.

1.1 Biotekniikan ja bioteknologian määritelmistä

Jos biotekniikalla tarkoitetaan esimerkiksi mikrobien hyödyntämistä elintarvikkeiden valmistuksessa, sen historia yltää jo muinoin keksittyihin hiivan käyttöön oluen ja leivontatuotteiden valmistuksessa sekä hapatemikrobien käyttöön hapanmaitotuotteiden, kuten juuston, valmistuksessa. Tietenkään biotekniikasta nykymuodossaan puhumattakaan biologisista tai molekyyli-tason tapahtumista ei ole tällöin ollut vielä mitään käsityksiä. Modernin biotekniikan alkuna pidetään yleensä DNA:n keinotekoisien muokkaamisen löytämistä (Rekombinantti-DNA -tekniikka) 1970-luvun alkupuolella. Bioteknisen teollisuuden juuret ovat samassa ajankohdassa. Yliopistojen rooli alan kehityksessä on ollut heti alusta alkaen erittäin merkittävä. (Tas-kinen, 2003, 5)

Biotekniikalle on monitieteisenä ja -alaisena alueena vaikea kehittää kattavaa ja yksikäsitteistä määritelmää. Kehittyneiden markkinatalousmaiden yhteistyöjärjestö OECD on pyrkinyt tekemään jäsenmaitaan varten synteetisistä erilaisista biotekniikan määritelmistä, joita eri organisaatiot ja toimijat ovat kehittäneet. Näitä eri määritelmiä on laaja kirjo, ja vuonna 1982 OECD:n julkaisemassa katsauksessa niitä on listattuna yksitoista kappaletta. Määritelmät vaihtelevat sen mukaan, millaisesta intressiryhmästä ja kokonaisnäkemyksistä on kyse. Kuten OECD:n katsauksessa todetaan, on tärkeää saada aikaan hyväksytty yhtenäinen määritelmä, jotta kansainvälinen vertailu ja keskustelu eri kansallisten toimijoiden kehityksestä on mahdollista. (Bull *et al.*, 1982, 18)

OECD:n rajaama määritelmä biotekniikalle on seuraavanlainen:

“The application of scientific and engineering principles to the proces-

sing of materials by biological agents to provide goods and services.”
(Bull *et al.*, 1982, 18)

Jossa määritellään biotekniikka luonnontieteellisten ja teknisten tieteiden periaatteiden hyödyntämiseksi tuotteiden ja palveluiden tuottamiseksi biologisten agenttien avulla.

Tässä tieteen ja tekniikan välineet viittaavat vahvasti mikrobiologiaan, biokemiaan sekä genetiikkaan. Tällöin esimerkiksi perinteistä kasvien ja eläinten jalostusta ei lueta kuuluvaksi biotekniikan alaan. (Bull *et al.*, 1982, 18). Biotekniikka ei siis ole erillinen tieteenala vaan joukko teknisiä ja tieteellisiä työkaluja (Taskinen, 2003, 6). Suomalaisessa tiede- ja teknologiapolitiikassa tämä näkyy niin, ettei biotekniikkaa yleensä kategorisoida omaksi alakseen esimerkiksi kaupparekisterissä (Luukkonen toim., 2004, 8).

Usein bioteknologia ja biotekniikka rinnastetaan synonyymeiksi. Tässä tutkimuksessa käsitteet kuitenkin erotetaan toisistaan. Biotekniikka nähdään rajatumpana käsitteenä, joka viittaa yksittäisiin tekniikoihin, prosesseihin sekä luonnontieteelliseen ja tekniseen tietoon. Tämän nojalla biotekniikan asiantuntijat ovat siis rajattavissa teknisten tieteiden ja luonnontieteiden soveltajiin. Bioteknologialla puolestaan viitataan tässä tutkimuksessa laajempaan teknologiseen verkostoon, jonka voidaan ajatella koostuvan toimijoista, teknisistä artefakteista sekä tiedoista ja prosesseista (esimerkiksi geenimanipulaatio).

Verkostomaista ajattelutapaa teknologisten järjestelmien hahmottamisessa on kuvannut esimerkiksi Michael Callon. Hän esittää teorian, jossa teknologia voidaan nähdä heterogeenisista elementeistä koostuvana muuttuvana toimijaverkostona. Tämä verkosto yhdistää toisiinsa elementtejä, jotka voivat perinteisestä sosiologias- ta poiketen olla ihmistoimijoiden lisäksi elottomia elementtejä. Callon esittää, että teknologian analyysissä on mielekästä tarkastella verkostoa yhdistämällä sosiologi- nen, teknologinen ja taloudellinen näkökulma, jolloin toimijajoukkoa on ajateltava varsin laajasti. (Callon, 1999)

Tässä tutkimuksessa nähdään bioteknologia laajassa mielessä teknologisenä järjes- telmänä, eli ei rajoituta pelkkään biotekniikan (eli tieteen, prosessien ja tekniikoi-

den) hyödyntämiseen liittyvään piiriin. Edellä oleva Callonin näkökulma otetaan siis perusteeaksi erilaisten toimijoiden kytkemiseksi bioteknologiakentän yleiseen tarkasteluun.

Tässä tutkimuksessa nojaututaan kohderyhmän osalta melko väljään bioteknologia-alan rajaukseen, koska analyysin luonteen vuoksi halutaan nähdä alan monitieteellisyys ja yhteistyö erilaisten toimijoiden välillä. Tähän liittyen halutaan kartoittaa yhteyksiä ja näistä syntyviä verkostoja, joita ei tiukasti rajatulla kohdeaineistolla voitaisi saada esille. Toisaalta tutkimuksessa pyritään tarkastelemaan kenttää yleisesti kartoittavalla tasolla; aineiston ja laskettujen tunnuslukujen pohjalta ei testata tutkimushypoteeseja. Tutkimuksen päätavoitteena on urapolkujen ja näihin liittyvien toimijaverkostojen tutkimiseen sopivien menetelmien kehittäminen.

Alkuperäisenä tarkoituksena oli tutkia suomalaisten bioteknologia-alan yritysjohtajien urapolkuja, mutta aineiston hankalan saatavuuden takia kohderyhmää laajennettiin. Lisäksi osoittautui mielenkiintoiseksi tutkia tutkimus- ja yrityskentän välisiä yhteyksiä. Mielekkäänä jatkotutkimuskohteena voisi kuitenkin olla rajatumman joukon, kuten bioteknologia-alan yliopistotutkijoiden tarkastelu.

Tutkimuksen kohderyhmä koostuu toimijoista, jotka toimialallaan sijoittuvat edellä läpikäydyn OECD:n bioteknologian määritelmän alueelle, mutta myös esimerkiksi rahoituspuolen toimijoita. Nämä ovat bioteknologia-alan tieteellisen tutkimuksen ja sen teknisen sovelluskentän piirin ulkopuolella, mutta ne voidaan ottaa mukaan analyysiin, koska ne liittyvät laajemmassa mielessä bioteknologia-alan teknologiseen verkostoon.

1.2 Bioteknologia Suomessa

Globaali kilpailu bioteknologia-alalla on voimakasta. Euroopan bioteknologiayritysten määrä on suurempi kuin Yhdysvalloissa, mutta täällä yritykset ja niiden kehittämien tuotteiden määrät ovat pienempiä. Euroopan komissio onkin tämän johdosta määritellyt strategian bioteknologia-alan kehittämiseksi. Strategia tähtää investointien tehokkaaseen ohjaamiseen menestyvien kaupallisten tuotteiden kehitykseen. Tämä vaikuttaa vahvasti Suomen bioteollisuuden tulevaisuuteen, ja stra-

tegiaehdotus on otettu positiivisesti vastaan. (Hassinen, 2002)

Suomessa on ollut ajatuksena rakentaa uusi teollisuuden tukijalka bioteknologiasta metalliteollisuuden, puunjalostuksen sekä tietotekniikan rinnalle. Esimerkiksi Turun alueen bioteknologia-osaamiskeskuksen keskeisenä ajatuksena on ollut visio alueen kehittämisestä alan markkinajohtajaksi. Tätä tarkoitusta varten tutkimuskeskuksia ja muuta infrastruktuuria on perustettu kampusalueen liepeille, jossa on myös palveluita ja muuta toiminnallisuutta. Tällä pyritään houkuttelemaan alueelle alan yrityksiä, osaamista ja kehittämään alan opetusta. Kunnianhimoisena tavoitteena on saada Turkuun bioteknologia-alalle 10000 uutta työpaikkaa. (Mantila, 2002)

Turun alueen kehityksessä merkittävintä roolia ei ole kuitenkaan vetänyt kansallinen tiede- ja teknologiapolitiikka, vaan teknologiakeskus BioCitystä muodostui niin sanottu rajaolio (*boundary object*), joka yhdisti erilaisten toimijoiden intressejä. Eri toimijat näkivät keskuksen omista erilaisista tarvenäkökulmistaan: teollisuudelle se edusti hyvää kontaktirajapintaa yliopistotutkimukseen, yliopistoille se toimi tutkimustoiminnan kehittäjänä, ja kaupungin kannalta se oli merkittävä taloudellinen toimija. BioCityn perustaminen teki siitä verkoston resurssikeskittymän, joihin toimijat olivat tavalla tai toisella vahvasti kiinnittyneet. (Höyrylä *et al.*, 2004)

Bioteknikan yritysten sovellusalueet voidaan jakaa kolmeen ryhmään: terveydenhuoltoon liittyviin, prosessiteollisuuden sovelluksiin ja palveluihin. Suurin ryhmä Suomessa on lääkkeisiin ja terveydenhuollon diagnostiikkaan liittyvät yritykset. Uuden biotekniikan yritysten määrä lähti selkeästi kasvamaan 1980-luvun lopulla, ja valtaosa yrityksistä on perustettu 90-luvulla ja sen jälkeen. (Luukkonen *toim.*, 2004, 9-11)

Monet bioteknologiayritykset ovat keskittyneet viiden bioteknologiakeskittymän ympärille Helsingissä, Turussa, Tampereella, Kuopiossa ja Oulussa. Tämä johtuu siitä, että paljon tutkimustietoa on saanut alkunsa yliopistoista sekä korkeakouluisista, ja yritykset jatkokehittävät sekä soveltavat näitä tuloksia. Suomalaiset bioyritykset ovat hyvin tutkimusorientoituneita, eikä tuotekehitys ole laajemmin tarkasteltuna vielä pitkällä. Toiset yritykset eivät etene tuotantotasolle asti, vaan keskittyvät alan tutkimukseen. Sovellusten kehitys vaatii suuria investointeja ja pitkiä kehitysaikoja, mutta menestykseen tarvitaan rahallisen panostuksen ja alan ydinosaamisen

lisäksi muidenkin alojen osaamista; näistä mainittakoon tietotekniikka, prosessitekniikka, talous, kuten markkinointi ja rahoitus, sekä patenttikysymyksiin liittyvä juridiikka. Euroopassa on myös paljon pohdittu bioteknologian eettisiä näkökohtia. Kaikki nämä erityyppiset vaatimukset alan teknologiapolitiikassa ja kaupallisissa sovelluksissa vaativat laajasti eri toimijoiden ja intressiryhmien vuoropuhelua ja yhteistyötä. (Hassinen, 2002)

1.3 Uratutkimuksista

Koska kyse on uratutkimuksesta, eli henkilöiden koulutus- ja työuratietoja käytetään primääriaineistona tutkimuksessa, tässä alaluvussa esitellään aluksi yleisesti uratutkimuksia, niissä käytettyjä lähestymistapoja ja tutkimusotteita.

Uratutkimuksia voidaan käyttää yleisesti urasuunnittelussa, organisaatiotutkimuksessa tai jonkin toimijajoukon, esimerkiksi suomalaisten toimitusjohtajien, tutkimuksessa. Tutkimusten tavoitteena voi olla yleisen mallin rakentaminen jostakin urailmiöstä tai uratyyppejen etsiminen jostakin joukosta.

Leo Ahlstedt on tutkinut suomalaisen liikkeenjohtajan uraa sekä kvalitatiivisella että kvantitatiivisella tutkimusotteella. Hän on pyrkinyt selvittämään kahta tekijää: liikkuvuutta ja liikkeenjohtajan erikoistumista. Tärkeimpänä tavoitteena on ollut oppia kuvaamaan liikkeenjohtajan urakehitystä, erityisesti sen rakennetta ja vaiheita. Ahlstedt on identifioinut tutkimuksessaan neljä liikkeenjohtajan uratyyppiä. Aineistona on ollut survey-kyselyn tulokset sekä tämän tukena uratyyppeihin pureutuva case-tutkimus, joka pohjautuu haastatteluihin. (Ahlstedt, 1978)

Uratutkimusta voidaan tehdä yhden organisaation sisällä. Tällainen tutkimus voi esimerkiksi antaa viitteitä siitä, kuinka organisaatiota tulisi kehittää rakenteellisesti, jos esimerkiksi ammatillisen kehityksen kannalta havaitaan jotain esteitä. Tutkimusten teoreettisena pohjana voi olla psykologisia tai sosiologisia teorioita, taikka sitten uraa tarkastellaan ulkopuolisen objektiivisen havainnoijan näkökulmasta.

Edellä selostettiin muutamia esimerkkejä kvantitatiivisista ja kvalitatiivisista näkökulmista uratutkimuksissa. Tässä tutkimuksessa kuvatuilla menetelmillä käsitellään

kvantitatiivista urapolkuaineistoa, jotta pystyttäisiin empiirisen aineiston tuella kuvaamaan suomalaisen bioteknologia-alan toimijoita sekä toimijoiden organisaatio-kontakteista muodostuvaa verkostoa.

1.4 Tutkimusraportin rakenne

Johdantoluvussa käsiteltiin biotekniikan ja bioteknologian määritelmiä tämän tutkimuksen näkökulmasta. Lisäksi esiteltiin lyhyesti suomalaista bioteknologia-alaa ja käytiin läpi urapolkututkimuksien kenttää.

Tutkimusraportin luvussa 2 selvitetään tämän urapolkututkimuksen keskeiset käsitteet sekä urapolun hahmottaminen kvantitatiivisessa analyysissä. Tämän jälkeen selostetaan tutkimuksen merkitys ja tavoitteet, sekä määritellään tutkimustehtävä.

Luvussa 3 siirrytään varsinaisen urapolkuanalyysin toteuttamiseen. Aineiston hankinnan ja arvioinnin jälkeen selvitetään tutkimuksen luotettavuutta ja tulosten yleistettävyyteen liittyviä näkökohtia. Seuraavaksi selvitetään aineiston kvantifointi sopivien muuttujien arvoiksi, jotta voidaan tehdä analysointi analyysiohjelmistolla. Sitten esitellään perustulokset aineistosta, eli kuvaillaan kohdejoukkoa perustietojen, kuten iän ja koulutuksen, osalta.

Seuraavaksi esitellään pääkomponenttianalyysin käyttö analyysityökaluna sekä tulokset tästä urapolkuaineistosta. Toinen menetelmä, eli verkostanalyysi, selostetaan tuloksineen luvun lopuksi.

Tutkimusta arvioidaan tarkemmin luvussa 4. Tässä kohdassa käydään läpi tutkimukselle asetetut tavoitteet ja arvioidaan miten hyvin ne saavutettiin. Lisäksi esitetään joitakin jatkotutkimusmahdollisuuksia käsillä olevan analyysin lähtökohdista.

Luku 2

Peruskäsitteet ja tutkimustehtävä

Tässä luvussa määritellään tutkimusongelma, sekä selvitetään tutkimuksen tärkeimmät käsitteet: urapolku ja ura-askel. Lisäksi hahmotellaan urapolun yleinen käsitteilymuoto tämän tutkimuksen analyysissä.

2.1 Käsitteet tutkimuksessa

Urapolku on jonkin henkilön erillisten toimien muodostama ketju, jota tietyllä hetkellä havainnoidaan. Urapolkua voidaan nimittää myös henkilön urahistoriaksi. Se kertoo ura-askeleissa henkilön siihenastisen työhistorian. (Ahlstedt, 1978, 29)

Urapolku on täsmällisempi käsite kuin ura, joka helposti mielletään kehitysprosessiksi, jossa edistytään koko ajan parempiin ja haastavampiin toimiin. Urapolku sisältää ajatuksen siirtymästä toimesta toiseen; työkokemus kehittyy kumulatiivisesti, ja sekä organisaatio että toimiala voivat vaihtua. Urapolku-käsitettä ja sen pohjalta rakennettavaa kvantisoitua aineistoa voidaan kenties kritisoida sillä perusteella, että verkostomaisten organisaatiomallien (Castells, 1996, 164-165) samoin kuin verkostomaisten työprojektien yleistyessä ei voida aina selkeästi määritellä yksikäsitteistä toimialaa, työnantajaa tai organisaatiota jossa työskennellään. Samoin perinteinen uramalli, jossa ollaan saman organisaation palveluksessa läpi koko elämän, on yhä harvinaisempi. Toisaalta verkostomaisten toimintamallien tutkimisessa urapolkua-

nalyysi voi toimia hyvinkin, jos vain aineistosta saadaan esille riittävän tarkkaa informaatiota yhteyksistä.

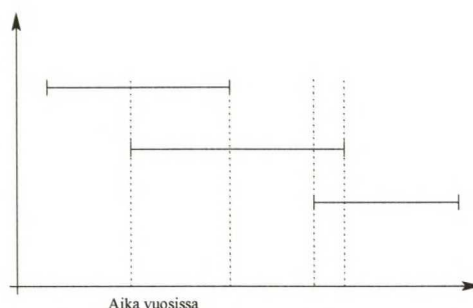
Urapolku erillisten toimien ketjuna on kuitenkin käyttökelpoinen tapa analysoida henkilön työhistoriaa, ja edellä kuvatut määrittelyn vaikeudet voidaan olettaa marginaalisiksi. Onkin syytä ajatella esimerkiksi ansioluetteloista muodostettavaa aineistoa vain osakuvauksena kokonaistyöhistoriasta. Se ei näin ollen kuvaa kaikkea työurasta, eikä sitä voi käyttää kaikissa tutkimuskysymyksissä. Kun tehdään johtopäätöksiä ja yleistyksiä urapolkuaineistosta, on varottava, ettei oleteta aineistoa kattavammaksi kuin se oikeastaan on. Jos esimerkiksi projektinomaisia työsuhteita on todellisuudessa paljon, mutta ne kuitenkin suurelta osin puuttuvat kerätyistä aineistosta, niin johtopäätökset toimijoiden välisestä yhteistyöstä eivät ole valideja.

Ura koostuu ura-askelista, jotka ovat yksittäisiä toimia. Ura-askeleella voidaan tarkoitaa kunkin hetkistä päätoimea, eli yksikäsitteistä toimea kullakin ajanhetkellä. (Ahlstedt, 1978, 29)

Esimerkki yksittäisestä ura-askeleesta voisi olla yrityksen toimitusjohtajan vakanssi. Ura-askeleen voidaan ajatella sisältävän tiedot toimesta, eli ainakin alkamisajankohdan, organisaation, toimen keston ja toimenkuvan. Ura-askelten kokonaismäärä kertoo suoraan uran kokonaisliikkuvuuden, eli kuinka monessa toimessa henkilö on tähänastisessa työhistoriassaan ollut.

Ketjumainen urapolun määritelmä voidaan ymmärtää niin, että koko urapolku koostuu toisiaan seuraavista erillisistä osista. Näin ei kuitenkaan todellisuudessa ole, vaan ura-askeleet voivat olla päällekkäisiä, kuten havainnollistavasta kuvasta 2.1 nähdään. Kuvitteellinen henkilö on työuransa alusta tutkimushetkeen saakka töissä yhteensä kolmessa eri toimessa, joista osa menee ajallisesti päällekkäin. Osa näistä toimista voi olla sivutoimia, jota esimerkin kuvaajassa voisi esittää keskimäinen jana.

Kuten aiemmin ura -käsitteen yhteydessä huomautettiin, urapolkua ei ajatella ammatillisessa kehityksessä koko ajan ylöspäin etenevää suuntauksena, jossa päästään koko ajan vastuullisempiin tehtäviin ja loppuvaiheessa toimitaan johtotehtävissä. Siirtyminen vähemmän vastuullisiin tehtäviin tai vastuutason pysyminen samana on mahdollista monessa tapauksessa. Tällainen arvottaminen ei ole edes kovin mie-

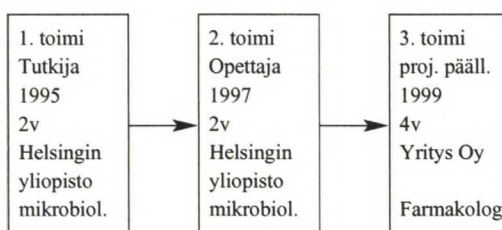


Kuva 2.1: Rinnakkaiset ura-askeleet

lekästä, kun ajatellaan urapolkuja yleisesti. Jos esimerkiksi yhtiön toimitusjohtaja perustaa oman yrityksen, ei voida yksiselitteisesti sanoa, miten tehtävän vastuullisuus ja merkittävyys on muuttunut aiempaan verrattuna.

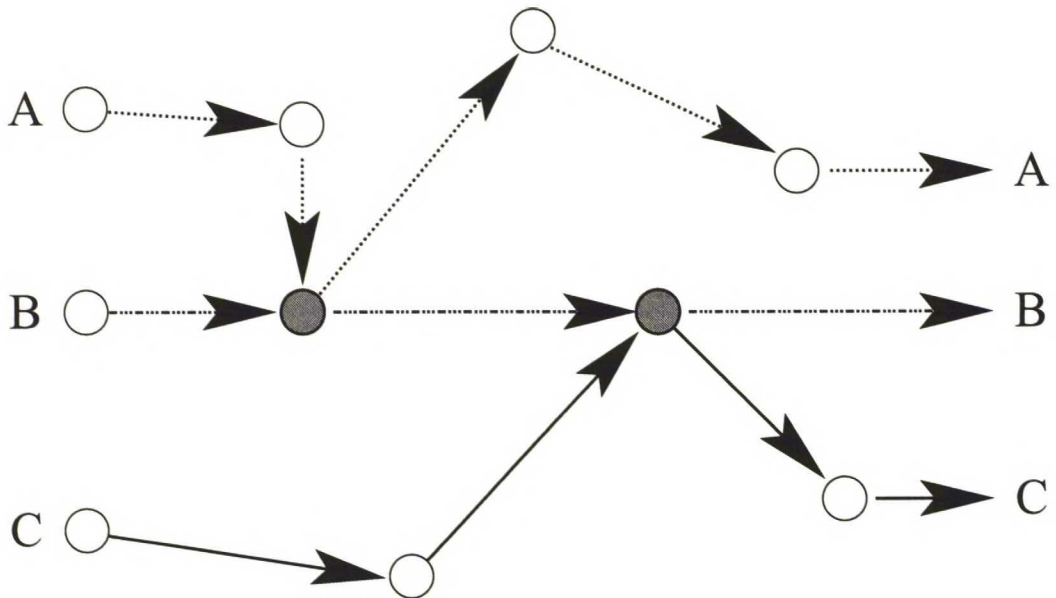
Tällainen arvotus on mielekästä vasta, jos rajaukset ovat selkeämmät. Esimerkiksi liikkeenjohtajan urapolulla on tietenkin kiinnostavaa tutkia minkälaisia vastuualueita johtaja on hoitanut (toimitusjohtajalla kokonaisvastuu yrityksen operatiivisista toiminnoista tai keskijohtoon kuuluvan toimijan osavastuu esimerkiksi markkinoinnista). Yleisesti ottaen mitä laajemmasta toimenkuvien joukosta on kyse, sitä hankalampaa on tehdä tämänkaltaisia arvoasteikkoja eri uramuuttujista.

Urapolku hahmotetaan tässä analyysissä päätoimien ura-askelten ketjuna (kuva 2.2). Jokainen ura-askel hahmotetaan tietueena, joka sisältää tiedot toimesta. Tärkeimmät tiedot ovat toimen alkamisajankohta, kesto, organisaatio ja toimenkuva.



Kuva 2.2: Ura-askelten ketju

Kun urapolkuaineisto on hahmotettu edellä kuvatulla tavalla, sitä voidaan visualisoida ja analysoida eri tavoin. Voidaan esimerkiksi ottaa lähtökohdaksi organisaatiot (tai niiden joukot), joissa eri toimijat ovat uransa aikana olleet. Tätä näkökulmaa pyrkii hahmottamaan kuva 2.3, jossa on kolme yksinkertaistettua urapolkua



Kuva 2.3: Organisaatioliikkuvuus

organisaatioliikkuvuuden näkökulmasta.

Graafissa nähdään kolmen henkilön - A, B ja C - urapolut. Kukin ympyrä kuvaa graafin solmua, joka on urapositio, tässä tapauksessa organisaatio. Voidaan ajatella, että useampi urapolku esittää organisaatiotopologiaa eli verkostoa, jossa henkilöt kulkevat.

Kukin henkilö aloittaa uransa toisistaan riippumatta jossakin organisaatiossa, jota kuvaavat ensimmäiset, vasemmanpuoleiset solmut graafissa. Henkilöiden A, B ja C urapolkujen keskivaiheella on nähtävillä kaksi yhteistä solmua. Analyysissä voidaan tutkia, kuinka monta urapolkua kulkee jonkin tietyn organisaatioluokan kautta. Jos frekvenssit ovat suuria, silloin tämä luokka mahdollisesti ilmentää merkittävää solmukohtaa kohderyhmän urakehityksessä. Jatkokysymyksenä voidaankin esittää, miksi ko. luokka on saanut niin suuria edustuksia aineistossa. Jos B:llä ja C:llä olisi useampia yhtymäkohtia urapoluillaan, polut saattaisivat edustaa samaa uratyyppejä joillakin kriteereillä ilmaistuna. Se voisi olla esimerkiksi siirtymä akateemiselta uralta yritysmaailmaan.

2.2 Tutkimuksen merkitys ja tavoitteet

Bioteknologia-alan toimijoiden urapoluista ei ole Suomessa merkittäviä tutkimuksia. Tämä tutkimus ottaa tarkastelun kohteeksi yksilönäkökulman eli yksittäisen bioteknologia-alan toimijan: minkälaisia uratyyppejä aineistosta nousee, ja minkälaista verkostoitumista toimijoiden välillä on.

Tämän tutkimuksen tavoitteena on kehittää kvantitatiivinen analyysimenetelmä urapolkujen tutkimiseen, ja testata menetelmää pienellä aineistolla, joka käsittää joukon suomalaisen bioteknologia-alan toimijoita. Hypoteeseja tässä tutkimuksessa ei aseteta empiirisen aineiston pienen koon ja vinoutumien takia. Suurin syy siihen ettei tilastollisia yleistysjä voidaan tehdä, on se, ettei otoksen poiminnassa ole voitu käyttää satunnaisotantaa johtuen syistä joita myöhemmin käsitellään. Opinnäytetyön puitteissa laajamittainen aineiston hankinta osoittautui liian työlääksi. Tutkimuksessa pyritäänkin rakentamaan menetelmällinen kehys jatkotutkimuksille, jossa laajemmasta ja kattavamasta otoksesta voidaan tehdä tilastollisia yleistysjä. Analyysikehys on tehty niin, että sitä voidaan hyödyntää laajemmissa tutkimuksissa käyttämällä tässä kuvattuja menetelmiä sekä työkaluja ja laajentamalla niitä tarpeen mukaan.

Empiirinen aineisto kvantifioidaan numeeriseen muotoon, ja analyysimenetelmän avulla siitä selvitetään uratyyppejen hahmoja, eli minkälaisia urien luokkia aineistosta voidaan löytää. Luokat identifioidaan klusteroimalla toimijat muuttujien arvojen suhteen. Lisäksi tarkastellaan toimijoiden samanaikaisista organisaatiopositiosta muodostuvaa verkostoa.

2.3 Tutkimustehtävä

Tutkimustehtävä on kolmiosainen:

1. Miten ansioluettelomateriaalista kvantifioitua dataa voidaan käyttää bioteknologia-alan urapolkujen ja niistä muodostuvien verkostojen tutkimisessa? Luodaan analyysimenetelmä, jolla tätä aineistoa voidaan käsitellä ja arvioi-

daan menetelmän toimivuus.

2. Hankitaan aineisto ja arvioidaan sen käytettävyyttä tutkimuksessa. Aineiston kvantifiointia varten kehitetään sopiva työkalu ja tämän jälkeen aineisto muunnetaan numeeriseen muotoon analyysiä varten.
3. Seuraavaksi tehdään varsinainen analyysi kvantifoidulla aineistolla, minkä jälkeen tarkastellaan, mitä havaintoja kohdejoukosta voidaan tehdä:
 - Minkälaiset urapolkuluokat tai -hahmot ovat tämän aineiston puitteissa vallitsevia, eli mikä on tyypillinen urapolku.
 - Minkälainen verkosto muodostuu, kun yhdistetään sellaiset toimijat, jotka ovat olleet yhtä aikaa samassa organisaatiossa.

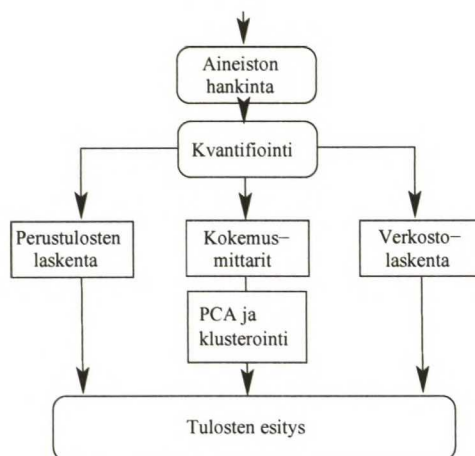
Luku 3

Analyysin toteutus

Tässä luvussa selostetaan kvantitatiivisen analyysin toteuttaminen, tarvittava matemaattinen teoria ja Matlab- ohjelmiston käyttö tulosten laskennassa. Aluksi kuvataan tutkimuksen kohderyhmä ja arvioidaan saatua aineistoa. Alaluvussa 3.2 tarkastellaan käsillä olevan aineiston luotettavuutta. Aineiston arviointi katsotaan kuuluvan yleisesti kehitetyn analyysikehikon osaksi, mistä syystä näkökohtia käsitellään hieman syvällisemmin, kuin juuri tämän aineiston kannalta olisi välttämätöntä. Tämän jälkeen kohdassa 3.3 käydään läpi urapolkuaineiston esikäsittely ja muuttujien analysointi vaihe vaiheelta. Esikäsitelty ja kvantifioitu data voidaan antaa analyysiohjelmiston syötteenä.

Alaluvussa 3.4 perustellaan käytetty analyysiohjelmo, ja esitetään analyysiproseduurien toteutus yleisellä tasolla. Alaluvussa 3.5 kuvaillaan aineistoa perustietojen osalta, jotta saadaan parempi käsitys kohderyhmästä. Alaluvuissa 3.7 ja 3.6 esitellään kaksi analysoinnin päämenetelmää ja niiden tulokset. Ensimmäisenä esitellään pääkomponenttianalyysi aineistossa esiintyvien uratyyppeiden hahmottamisen apuvälineenä ja toisena menetelmänä verkostanalyysi toimijoiden organisaatiokytkösten selvittämisessä. Saatua analyysituloksia arvioidaan luvussa 4.

Analyysin vaiheittainen toteutus on esitetty kuvan 3.1 vuokaaviossa:



Kuva 3.1: Analyysin toteutus

3.1 Aineisto

Alkuperäisenä diplomityön tutkimuskohteena oli bioteknologia-alan yritysten avain-toimijoiden (lähinnä toimitusjohtajien) ammatillisten kohtaamispisteiden ja ura-tyyppien tutkiminen. Koska kävi niin, että riittävää aineistomäärää oli hyvin han-kala saada, päädyttiin laajentamaan kohderyhmää koskemaan myös julkisen tutki-muksen ja rahoituspuolen toimijoita. Yritysten kartoitus tehtiin Suomen bioteolli-suus ry:n jäsenyritysluettelon pohjalta. Suomen yliopistot käytiin systemaattisesti läpi, ja otettiin mukaan joukko bioteknologia-alan professoreita ja tutkijoita. Myös suurimmat alan rahoittajaorganisaatiot otettiin tarkasteluun mukaan, vaikkakin näi-den toimijoiden osuus jäi lopulta aineistossa hyvin pieneksi.

Toisaalta kohderyhmän laajentaminen herätti kiinnostavan kysymyksen biotekno-logia-alan verkostoluonteen tutkimisesta, toisaalta taas kohderyhmästä tuli varsin heterogeeninen, mikä aiheuttaa tiettyjä vinoutumia analyysin tuloksiin. On esimer-kiksi varsin selvää, että yliopistotutkijoiden ja yritystoimijoiden urapoluista tulee ainakin jossain määrin omat tyyppinsä. Mutta kun tarkoituksena on tarkastella ver-kostoitumista ja mahdollisia tiedon integraation solmukohtia toimijoiden välillä, niin tällöin erilaisista toimijoista koostuva kohderyhmä on mielekäs.

Tutkimuksen primääriaineistoksi kerättiin tietoja bioteknologia-alalla toimivien johto-sekä esimiesasemassa ja asiantuntijatehtävissä olevien henkilöiden koulutuksesta ja

urakehityksestä. Toimijat ovat yrityksistä, jotka tekevät alan tutkimusta, konsultointia ja tuotekehitystä sekä julkisen tutkimuksen piiristä, kuten yliopistoista ja tutkimuslaitoksista. Mukana on myös muutamia rahoituspuolen toimijoita. Joukkoa on tarkoitus tarkastella kokonaisuutena, koska bioteknologia-ala on luonteeltaan monitieteistä ja organisaatioiden välinen yhteistyö on yleistä. Esimerkiksi lääkekehityksessä tehdään paljon yhteistyötä pienten ja suurten yritysten välillä. (Luukkonen toim., 2004, 53-57). Samoin yhteistyö yliopistomaailman ja yritysten välillä on tyypillistä, koska useinkaan yrityksillä ei ole mahdollisuuksia uusimman tieteellisen tiedon saamiseen omin voimin. Suhteet julkiseen tutkimusmaailmaan pohjautuvat toimijoiden välisiin sosiaalisiin verkostoihin epäformaalimmalla tasolla. (Luukkonen toim., 2004, 60-61)

Tutkimuksessa ei tarkastella henkilön taustojen ja persoonallisuustekijöiden merkitystä uran kehityksessä; tällaiseen tutkimukseen tarvittaisiin toisenlaista empiiristä materiaalia. Tässä tutkimuksessa tarkastellaan suomalaisten bioteknologia-alan toimijoiden uria kartoittamalla minkälaisia yhteisiä piirteitä niistä löytyy, ja onko havaittavissa samanaikaisia kohtaamispisteitä ammatillisella kehityspolulla. Muodollista uraa, eli toimien ketjua, tarkastellaan sellaisena kuin se on, ja pyritään löytämään yhteisiä tekijöitä ja ilmiöitä.

Tutkittavilta henkilöiltä pyydettiin ansioluettelo tai vaihtoehtoisesti toivottiin heidän ilmoittavan vastaavat ura- ja koulutustiedot muulla tavoin. Kyselyt tehtiin sähköpostitse ja puhelimitse. Myöhemmin lähestyttiin kirjeitse niitä, jotka eivät vastanneet. Joidenkin henkilöiden tiedot kysyttiin lyhyessä puhelinhaastattelussa.

Ansioluetteloiden käyttäminen kvantitatiivisen tutkimuksen aineistona on käsitteelyn kannalta työläämpää verrattuna lomaketyyppiseen survey-kyselyyn. Esimerkiksi survey-tyyppisestä urapolkuihin liittyvän aineiston keruusta on Leo Ahlstedtin tutkimus, joka käsittelee liikkeenjohtajien urapolkuja (Ahlstedt, 1978). Kysymyslomakkeen avulla voidaan valmiiksi strukturoida kysymykset niin, että vastaaja tavallaan esiluokittelee tietonsa. Tällöin varsinaiseen aineiston kvantifointiin, eli numeeriseen muotoon muuntamiseen, ei vaadita niin suurta työmäärää. Tässä tutkimuksessa päätettiin kuitenkin hyödyntää ansioluettelopohjaista materiaalia, mikä tietyistä eduista huolimatta aiheuttaa hieman suuremman vaivan tarvittavien muut-

tujien konstruoinnissa ja niiden arvojen luokittelussa. Tärkeä peruste ansioluettelo-materiaalin käyttöön on henkilöiden kiireisyys: kyselylomakkeeseen vastaaminen vie aikaa, ja kynnys kyselyn täyttämiseksi on suuri, jos henkilö ei katso, että hän tai hänen organisaationsa hyötyisivät tutkimuksesta riittävästi. Yleisenä ongelmana on siis vastaajan sitouttaminen tutkimukseen niin, että hän kokee asian riittävän tärkeänä. Ansioluettelon lähettäminen on vaivattomampaa kuin survey-kyselyyn vastaaminen, ja todennäköisesti tällöin ei tarvita niin suurta sitouttamisen astetta. Lisäksi ansioluettelot ovat varsin vakiintuneita muodoltaan, jolloin tarpeellisen koulutus- ja urainformaation selvittäminen niistä on verrattain helppoa (ks. taulu 3.1). Rajallisen luonteensa takia tässä tutkimuksessa ei ollut riittävästi resursseja laajan ja systemaattisen otoskehyksen tekemiseen.

Joitakin aineiston luokitukseen liittyviä ongelmia juontuu siitä, ettei tutkimukseen osallistuja pääse itse luokittelemaan vastauksiaan, vaan sen tekee tutkimuksen suorittaja. Tällöin kvantifiointivaiheen ongelmia voi syntyä esimerkiksi tehtävänimikkeiden sisällön kanssa. Ei esimerkiksi ole selkeää tietoa siitä, mikä on toimen sisältämän vastuun määrä, vaan sama toiminimike voi sisältää täysin erilaisen vastuun määrän erikokoisissa organisaatioissa. Kysymyslomakkeen käyttö ei ollut mielekäästä myöskään sen takia, että tämän tason toimijoita ei ole Suomessa kovin suurta

1. **Henkilötiedot**

Nimi, syntymävuosi

2. **Koulutus (korkein suoritettu tutkinto)**

Lääketieteen tohtori, Helsingin yliopisto, 1985

3. **Työpositiot (organisaatio, tehtävä ja ajankohta)**

Tutkimusassistentti, Helsingin yliopisto,

lääketieteellinen tiedekunta, 1980-1985

Tutkija, Kansanterveyslaitos, infektiaudit, 1985-1990

...

Taulukko 3.1: Fiktiivinen esimerkki ansioluettelosta, olennaiset tiedot

joukkoa, ja näinollen riski liian suuresta vastauskadosta olisi ollut huomattava.

Aineiston luokituksen ja jäsennyksen tarkoitus on helpottaa sen muuntamista kvantitatiiviseen muotoon, jotta voidaan laskea tilastollisia tunnuslukuja kohdejoukosta.

Aineisto jakautuu kolmeen osaan: yritykset, jotka tekevät alan tutkimusta, konsultointia ja tuotekehitystä; julkinen tutkimus, eli yliopistot ja tutkimuslaitokset sekä muutamien alan rahoittajat.

Tutkimuksen kohderyhmän ensimmäisenä osajoukkona on suomalaisten bioteknologia-alan yritysten operatiivista johtoa sekä asiantuntijoita. Koska aineiston saaminen osoittautui ennakoitua vaikeammaksi, tämä joukko jäi varsin pieneksi, ja saadut tulokset ovat tästä johtuen huonosti yleistettävissä.

Yrityksineiden kohdalla kohderyhmää hahmotetaan seuraavalla tavalla: Yritys on sosiaalinen järjestelmä, joka voidaan yksinkertaistaen kuvata tasoina. Tasot ovat operatiivinen johto, esimiestaso ja suoritettava taso. Lisäksi esimerkiksi asiantuntijat voivat asettua sekä suoritettavalle tasolle että esimies- ja johtotehtäviin. Tämä jäsenyys helpottaa kohderyhmän rajaamista: siis yrityspuolen avainhenkilöt tutkimuksen kannalta ovat toimitusjohtajia, tutkimusjohtajia ja vastaavia operatiivisen johdon henkilöitä, joilla on kokonaisvastuu tai osavastuu yrityksen toiminnoista. Lisäksennä ryhmään kuuluvat myös asiantuntijat.

Esimerkiksi toimitusjohtajalla on kokonaisvastuu yrityksen toiminnasta, joten on selvää, että hänen roolinsa on merkittävä yrityksen, ja tietyssä määrin koko alan kannalta. Muilla johtotason toimijoilla on osavastuu esimerkiksi yrityksen taloushallinnosta (talousjohtaja) tai tutkimuksesta ja tuotekehityksestä (t&k-johtaja). Tutkimuksen lähdeaineistossa on tämän merkittävyydestarkastelun perusteella mukana johtotason henkilöitä sekä muutamia asiantuntijoita.

Yrityspuolelta tehdyn näytteen otoskoko on 86, josta saatiin vastaukset 26:lta, eli vastausprosentti oli 32,1%. Näistä vastauksista jouduttiin kuitenkin vielä hylkäämään kaksi, koska niiden tiedot olivat hyvin ylimalkaisia, eivätkä ne sisältäneet riittävästi informaatiota analyysin tarpeisiin.

Yrityksineistossa on nähtävissä melko heikko vastausprosentti, joka osaltaan johtuu yritysjohdon kiireisyydestä. Lisäksi eräs vastauskatoon vaikuttava tekijä saat-

toi olla se, että ansioluetteloa ei voi palauttaa samassa mielessä nimettömänä kuin vastauksia kysymyslomakkeeseen. Toisilla henkilöillä saattoi olla periaatteellinen tapa kieltäytyä lähettämästä ansioluettelotietojaan. Yhden ryhmän muodostavat ne, joilla ei ollut ajan tasalla olevaa ansioluetteloa ja jotka eivät sitä tutkimusta varten alkaneet erikseen koostamaan. Puhelinhaastattelun tai sähköpostitse tehtävän kyselyn avulla pyrittiin näitä ongelmia korjaamaan, mutta tällä ei saatu merkittävää parannusta tilanteeseen.

Tutkimusetiikan ja tietojen luottamuksellisen käsittelyn takia ei aineistoa esitellä tässä niin yksityiskohtaisesti, että olisi riskinä yksilöidä henkilöitä. Tämän takia myöskään listaa yrityksistä ei tässä esitetä, koska tästä voitaisiin esimerkiksi päätellä, että tietyn yrityksen toimitusjohtaja saattaa olla aineistossa mukana. Analyysin tulokset esitetään tunnuslukuina. Yksittäistä henkilöä ei siis tutkimuksesta voida tunnistaa. (Tutkimusetiikasta; ks. esim. Alkula *et al.* (1994) s. 295-296).

Toinen kohderyhmän osa on bioteknologia-alan julkinen tutkimuskenttä eli yliopistot, korkeakoulut, ja muutamat julkiset tutkimuslaitokset. Tähän ryhmään on siis valittu kyseessä olevien organisaatioiden sellaiset esimiesasemassa olevat henkilöt, jotka ovat lähimpänä bioteknologia-alan toimintoja. Esimerkiksi yliopiston kemian laitoksella saattaa olla biokemian osasto, jolla tehdään bioteknologista tutkimusta. Tällöin organisaatiosta on valittu biokemian osaston esimies, eikä koko kemian osaston johtajaa. Tämä on tärkeää sen takia, että keskitytään tutkimaan nimenomaan bioteknologia-alalla toimivien ihmisten urakehitystä ja ammatillisia kohtauspisteitä. Rajausta on useissa tapauksissa kuitenkin varsin hankalaa, ja hyvä otos vaatisi huolellisen tarkastelun.

Viimeisenä kohderyhmän osana ovat rahoittajatahot. Bioteknologian rahoittajaorganisaatioita on Suomessa vähän. Tässä tutkimuksessa on mukana joitakin henkilöitä tältä alueelta, mutta tämä joukko on niin pieni, ettei sitä ole mielekäästä käsitellä tutkimuksessa erillisenä joukkona. Tilastollisessa käsittelyssä joukko toimii vain kokonaisuuden osana.

Julkiselta puolelta aineistoa kerättiin 92:lta toimijalta organisaatioista, jotka olivat yliopistoja, korkeakouluja ja julkisia tutkimusorganisaatioita. Vastaukset saatiin 42:lta. Vastausprosentti tässä ryhmässä oli 45,7% eli selvästi korkeampi kuin

yritysten kohdalla.

Seuraavaksi käsitellään yleisesti ansioluettelomateriaalin käyttöä tutkimusmateriaalina. Ansioluettelot ovat melko standardimuotoisia, eli tiedot koulutuksesta sekä työurasta ovat selkeästi nähtävissä. Yleensä mukana on ainakin korkein suoritettu tutkinto, koulutusala, sekä ammattipositiodien osalta organisaatio, toimi ja toimesaoloaika. Yksityiskohtien osalta ilmenee kuitenkin melko paljon ongelmia.

Ansioluettelo edustaa henkilön muodollista urapolkua. Kuten aiemmin selvitettiin, urapolku kuvaa henkilön tähänastista työhistoriaa yksittäisten ura-askelten muodossa. Urapolku käsitteenä on yksinkertaistus, koska henkilö voi toimia useassa toimessa yhtäaikaaisesti, eikä varsinaista päätoimea voi aina selkeästi määritellä. Esimerkiksi yliopiston professori voi samanaikaisesti toimia toisessa virassa tai yritysmaailman palveluksessa. Yksikäsitteistä päätoimen määritelmää ei voidakaan tehdä, vaan se on usein subjektiivinen näkemys. Tässä tutkimuksessa pyritään kvantifioimaan aineisto niin, että merkittävät rinnakkaiset toimet saadaan mahdollisimman täydellisinä esikäsiteltyyn dataan mukaan kuitenkin korostaen sitä, että kyseessä on päätoimien ketju. Toisaalta kaikkia projektimuotoisia tai selvästi toissijaisia toimia ei dataan sisällytetä, vaikka niiden merkitys joissakin tapauksissa voi olla merkittävä henkilön urakehityksessä hänen omasta mielestään, tai objektiivisestikin henkilön urapolkua ja lisätietoja analysoimalla. Tässä tutkimuksessa ei voida kuitenkaan mennä näin syvälle, koska lisätietoina tarvittaisiin oikeastaan henkilön haastattelu, jossa selvitettäisiin työhistoria perusteellisemmin. Lisäksi mahdollisia jatkotutkimuksia varten on huomioitava, että niissä tarkasteltavat tutkimuskysymykset asettavat tiettyjä vaatimuksia aineistolle. Hypoteesit ja tutkimusongelma määrittävät aineiston rajauksen ja vaadittavan kokonaistarkkuuden, jotta saataisiin valideja ja relevantteja tuloksia.

Edellä kuvattujen ongelmien takia joudutaan tyytymään päätoimen osalta subjektiiviseen näkemykseen, ja tehdään oletus, että päätoimet ovat analyysin kannalta merkittävimpiä, kun tarkastellaan uratyyppejä ja organisaatioyhtymäkohtia. Datan esikäsitteily on suhteessa siihen, minkälaista analyysiä ollaan tekemässä, joten sivutoimi on otettu tapauskohtaisesti mukaan, jos sen katsotaan olevan riittävän merkittävä uratyyppiin vaikuttavana tekijänä, tai muuten tärkeästä syystä. Pääosin sivutoi-

met on kuitenkin rajattu analyysistä pois. Tämä tuo mahdollisesti pientä vinoutumaa analyysiin. Kuitenkin kaikkien sivutoimien mukaanotto vaatisi melko täydellisiä uratietoja, mikä ei lähtökohtaisesti ansioluettelomuotoisessa datassa ole järkevä oletus. Tämä lisäisi myöskin työmäärää tämäntasoisessa tutkimuksessa varsin paljon suhteessa arvioituun hyötyyn.

Urapolun pituus lasketaan ura-askeleen määrässä. Esimerkiksi jos henkilö on ollut viidessä toimessa tähänastisen uransa aikana, urapolun pituus on viisi, riippumatta toimien ajallisesta kestosta. Urapolun pituutta käytetään henkilön kokonaisuraliikkuvuuden mittarina. Urapolun pituus aiheuttaa joitakin ongelmia ansioluettelomateriaalissa. Toiset vastaajat ovat saattaneet sisällyttää ansioluettelonsa hyvin tarkkaan koko työhistorian, sisältäen lyhyetkin työsuhteet ja toimen nimikkeiden muutokset. Toiset taas ovat kuvanneet työhistorian ylimalkaisemmin, tärkeimpien toimien luettelona. Tästä johtuen voi uraliikkuvuuden kohdalla olla pientä epätarkkuutta.

Uraliikkuvuuteen liittyvää ongelmaa voidaan pienentää jättämällä kirjaamatta aineistoon hyvin lyhyet työsuhteet, joilla voidaan katsoa olevan vain marginaalista merkitystä henkilön ammatilliselle erikoistumiselle ja urakehitykselle. Pääsääntöisesti alle puolen vuoden pituisia työjaksoja ei ole otettu analyysiin mukaan.

Nyt on siis rajattu aineisto niin, että tutkimuksissa työuraa ajatellaan päätoimien ketjuna, mutta jätetään aivan lyhyet toimet pois. Sivutoimina pidetään, kuten yleisesti ajatellaan, esimerkiksi osakeyhtiön hallituksen jäsenyyttä, vaikka sen rooli voikin olla henkilön omasta mielestä merkittävä. Ansioluetteloihin merkitään uratiedot yleensä nimenomaan päätoimien osalta. Se, onko sivutoimia merkitty, vaihtelee paljon ja riippuu miten merkittäväksi henkilö sivutoimen katsoo. Usein on myös niin, että henkilö on räätälöinyt ansioluettelon tiettyä työnhakutilannetta varten. Tällöin siinä luonnollisesti esiintyvät vahvimmin ne työpaikat, jotka ovat sen hetkisen työnhakutilanteen kannalta olennaisimpia. Voidaan kuitenkin olettaa, että henkilö on ansioluetteloa laatiessaan kuvannut merkittävimmät positiot esimerkiksi yleistä ammatillista kehitystä, johtamiskokemustaan ja muita vastaavia tekijöitä urahistoriastaan silmälläpitäen, vaikkei hän olisi kokonaisuudessaan kaikkia toimia siihen merkinnyt.

Luokitusongelmia saattavat aiheuttaa myös projektiluontoiset työsuhteet, joita on tehty verkostomaisesti. Tällaisessa verkostoprojektissa saattaa olla toimijoita useista organisaatioista. Henkilö saattaa olla verkostoprojektissa vaikkapa tutkijana, asiantuntijana tai koordinaattorina, ja verkosto saattaa olla toiminnassa vain tämän yhden projektin työstämisen ajan. Tällaisia verkostomaisia työprojekteja ei välttämättä merkitä ansioluetteloon, elleivät ne ole henkilön urakehityksen tai jonkin työnhakutilanteen kannalta erittäin tärkeitä. Jos taas sellainen on merkitty, tämän ura-askleen tiedot voivat olla hyvin epäselviä. Esimerkiksi toimen vastuut ja toimenkuva voivat olla hankalasti löydettäviä tietoja.

Projektiluontoisuus ja verkostomaisuus voivat aiheuttaa myös päällekkäisyyttä – henkilö saattaa tehdä useita projekteja eri organisaatioissa samanaikaisesti. Tällöin kysymys, mikä on päätoimi, hämärtyy. Tällaisissa tapauksissa on otettu pidempi ja merkittävänluontoisempi toimi päätoimeksi, jolloin mahdollinen virhe pyritään identifioimaan ja tekemään systemaattiseksi, jolloin sitä on helpompi hallita. Näitä tapauksia ei ole kuitenkaan merkittävästi, joten ongelma ei muodostune kovin suureksi.

Kuten aiemmin todettiin, ansioluettelot räätälöidään usein työnhakutilanteeseen sopivaksi, jolloin joitakin toimia saatetaan jättää pois. Henkilö ei ole tällöin katsonut näitä relevanteiksi, tai ovat jopa rasitteena ansioluettelossa toimea haettaessa. Tällaisen ansioluettelon saaminen tutkimukseen aiheuttaa myös pientä vinoutumaa. Kyselylomake korjaisi tätä ongelmaa jonkin verran, mutta ei välttämättä kokonaan. Jos kyselylomakkeessa pyydetäisiin vastaajaa kertomaan tietyillä luokituksilla vaikkapa kymmenestä edellisestä työsuhteestaan, näiden valinta ja luokitteluvastuu vain siirtyisi vastaajalle. Lisäksi pitkien työurien kohdalla kymmenen viimeisintä työpositiota ei välttämättä sisältäisi joitakin uran tärkeitä käännekohtia.

Yhteenvedona voidaan todeta, että ansioluetteloiden käyttö aineistona on haastavaa, vaikka niiden kerääminen aineistoksi saattaa olla melko vaivatonta. Ongelmia tulee lähinnä erilaisten tarkkuuksien ja merkintöjen kirjosta, sekä siitä mitkä toimet ylipäätään on merkitty ansioluetteloon.

3.2 Validiteetti ja reliabiliteetti

Tässä vaiheessa on järkevää tarkastella tutkimuksen ja aineiston luotettavuutta. Tutkimuksessa yhtenä tavoitteena on kuvailla saatujen tulosten avulla kohdejoukkoa, mutta yleistyksiä perusjoukkoon ei tehdä aineiston puutteiden ja pienen koon takia. Luotettavuusnäkökohdat ovat kuitenkin yleisesti tärkeitä, ja mahdollisessa jatkotutkimuksessa olisi tehtävä perusteellisempi analyysi siitä, mikä on aineiston luotettavuus sekä yleistettävyyys perusjoukkoon.

Analyysin kokonaisluotettavuutta ilmentävät kaksi mittaria: reliabelius ja validius. Jos jokin muuttuja eli käsitteen operationalisointi ei vastaakaan reaalia maailmaa täysin, vaikuttaa se tuloksien validiteettiin. Tällöin mittaustulos on tietyn verran virheellinen toistuvasti ja systemaattisesti. Tällainen tilanne saattaa johtua mittarin tai muuttujan operationalisoinnin ongelmista mutta myös dataongelmista, jos jotakin olennaista on jäänyt pois aineistosta. Jos taas virhe tapahtuu satunnaisarvoisesti suuntaan taikka toiseen, puhutaan reliabiliteetin alenemisesta. Tyypillinen esimerkki alentuneesta reliabiliteetista on tulosten tallennuksessa tapahtuva virhe. Nämä virheet ovat luonteeltaan sellaisia, että niiden odotusarvo intervalliasteikoisella muuttujalla on nolla, eli virhettä tapahtuu keskimäärin yhtä paljon kumpaankin suuntaan. (Alkula *et al.*, 1994, 89-94)

Tutkimuksessa määritellyt käsitteet on operationalisoitava sopivien mittareiden avulla. Jos mitataan henkilön kokonaisliikkuvuutta, selkeä mittari tähän on henkilön urapolulla ilmenevien toimien lukumäärä. Systemaattista virhettä ei esiinny, jos urapolun alku määritellään selkeästi, ja toimet rajataan niin, että vain tietyn ajallisen keston (esimerkiksi yli puoli vuotta) kestävät toimet otetaan mukaan. Jos päätoimet ja sivutoimet ovat aineistossa limittäin ja sekaisin, ei voida välttämättä muodostaa yksikäsitteistä päätoimista muodostuvaa urapolkua. Jos näiden katsotaan aiheuttavan systemaattista virhettä, eli validius kärsii, on pohdittava, kuinka näitä virhetekijöitä voitaisiin minimoida. Systemaattisen virheen toinen lähde on aineiston kvantifiointivaihe – miten luokitukset valitaan ja perustellaan.

Aineiston keruussa, käsittelyssä ja analyysissä tapahtuvat satunnaiset virheet alenavat tulosten reliabeliutta. (Alkula *et al.*, 1994, 94) Satunnaisvirheiden lähde voi

siis olla jo ansioluetteloissa, mutta suuri mahdollisuus virheisiin on myös kvantifointivaiheessa.

3.2.1 Yleistettävyyys

Yleistäminen on usein kvantitatiivisessa tutkimuksessa varsin oleellinen kysymys. Halutaan tietää, miten hyvin pienemmästä joukosta tehdyt päätelmät, esimerkiksi keskiarvot joistakin muuttujista, pätevät suuremmassa joukossa. Yhteiskuntatieteissä otantaan liittyvät ongelmat ovat yleisiä ja usein melko monitahoisia. Tilastotieteen otantateoriassa pyritään löytämään vastauksia siihen, miten otanta saadaan tehtyä niin, että tulokset ovat luotettavia. Tässä alaluvussa käsitellään otantaan liittyviä ongelmia yleisesti ja toisaalta liittyen käsillä olevan tutkimukseen.

Tässä tutkimuksessa pyritään löytämään kohderyhmästä yhteisiä tekijöitä ja ilmiöitä. Näiden yleistettävyyys kohderyhmään laajemmin on ongelmallista. Ensinnäkin saatu aineisto-otos ei edusta satunnaisotosta perusjoukosta. Kun vastausprosentti on hyvin alhainen, joukko on mahdollisesti valikoitunut tiettyjen kriteerien mukaan. Ne, jotka vastasivat ja jättivät vastaamatta, eivät välttämättä ole satunnaisesti ryhmässä esiintyviä, vaan aineistoon tulee systemaattista virhettä. Vastaamatta jättäminen voi liittyä vaikkapa organisaation suhtautumiseen ulkopuolelta tuleviin kyselyihin. Tällainen suhtautumistapa voisi olla kategorinen kieltäytyminen tutkimuskyselyistä, kun halutaan kaikki panostustus organisaation operatiivisiin tehtäviin tai halu suojella vahvasti organisaation salaisuuksia, olivat ne sitten johtamiseen tai tuotantoon liittyviä. Tällaiset organisaatiot jäävät aineistosta pois, ja se voi aiheuttaa tiettyä vinoutumaa. Samoin ne, jotka jättävät vastaamatta kiireiden takia, eivät todennäköisesti jakaudu satunnaisesti, vaan tämä joukko saattaa sisältää erittäin merkittävän osajoukon, jonka puuttuminen otoksesta näkyy selkeästi tuloksissa.

3.2.2 Otanta

Perusjoukko eli populaatio on se joukko, jota halutaan tutkia, esimerkiksi kaikki suomalaiset bioteknologia-alan tutkijat tietyllä tavalla rajattuna. Tutkimalla kaik-

ki perusjoukon alkioit saadaan tietysti luotettavin empiirinen kuva todellisuudesta, mutta tähän ei tietenkään ole yleensä mahdollisuutta. Kun poimitaan perusjoukosta pienempi joukko jollakin satunnaismenetelmällä, puhutaan otoksesta. Satunnaismenetelmällä tässä tarkoitetaan sellaista tekniikkaa, jonka avulla voidaan poimia perusjoukosta jokainen alkio satunnaisesti toisistaan riippumatta. (Milton ja Arnold, 1995, 190-193)

Kun kehitetään mahdollista listaa perusjoukon jäsenistä, ei välttämättä saada kehykseen kaikkia alkioita. Tällöin on tärkeää arvioida, kuinka suuri joukko perusjoukosta on tässä kehyksessä, josta lähdetään otosta poimimaan. Eli on selvitettävä, onko joukosta pois jokin merkittävä osajoukko, vai voidaanko puuttuva osa arvioida niin pieneksi, ettei sillä ole juurikaan vaikutusta tuloksiin.

Kun ei käytetä satunnaispoimintaa, puhutaan näytteen ottamisesta perusjoukosta. Näytteen ottaminen voidaan tehdä niin, että pyritään saamaan näytteeseen perusjoukon ominaisuudet samassa suhteessa. Koska ei ole mitään takeita siitä, että jokainen näytteen alkio on perusjoukon satunnainen valinta, ei voida tehdä tilastollisia päätelmiä tuloksista, kuten laskea luottamusvälejä tuloksista. Luotettavuudesta ei ole siis mitään matemaattisia mittareita, ja on tyydyttävä vain kvalitatiiviseen arviointiin. (Alkula *et al.*, 1994, 106-107)

Näyte voidaan poimia monella tavalla: Voidaan pyytää vapaaehtoisia ilmoittautumaan tutkimukseen, tai tutkija voi itse poimia jonkin kriteerin perusteella tutkittavat. Eräs tapa on poimia tutkittavat yksinkertaisesti sen mukaan, miten he ovat saatavilla. Toinen tapa on käyttää lumipallomenetelmää, jota käytetään, kun listaa perusjoukosta on mahdotonta tai hankala muodostaa. Lumipallomenetelmässä jokaisen tutkittavan annetaan itse kertoa, ketkä olisi mielekästä ottaa seuraavaksi tutkimukseen. (Fink, 1995, 32-34)

Kun poimintamenetelmä on määritelty, olkoon se sitten satunnaisotannon hyväksikäyttö taikka lumipallomenetelmän soveltaminen, voidaan lähestyä henkilöitä haastattelun tai survey-kyselyn puitteissa taikka voidaan pyytää henkilöiltä antamaan jotain materiaalia. Tässä tutkimuksessa käytetään näytteen ottamista sen mukaan, miten aineisto on saatavilla.

3.2.3 Tämän tutkimuksen erityiskysymyksiä

Tämä tutkimus on luonteeltaan metodien soveltamista kokeileva, joten ei pyritä tekemään yleistyksiä näytteestä perusjoukkoon. Tässä kuitenkin esitetään jatkotutkimuksia silmälläpitäen otoksen muodostamiseen ja yleistysten tekoon liittyviä pohdintoja, koska ne liittyvät käsillä olevaan analyysitapaan keskeisesti.

Jos halutaan tehdä yleistettäviä päätelmiä suomalaisen bioteknologiakentän jostakin ryhmästä, on ensinnäkin määriteltävä perusjoukko huolellisesti, tämän jälkeen on tehtävä otoskehikko ja poimittava otos perusjoukosta.

Kuten Fink toteaa, hyvä otos on miniatyyriversio koko populaatiosta, samanlainen, mutta pienempi. Hyvä otos on edustava, eli otoksen ominaisuuksien jakaumien on oltava samanlaiset suhteessa perusjoukkoon. Yleensä määritellään ensin perusjoukko laajasti, ja sitten esitetään inkluusio- ja ekskluusiosäännöt, joiden pohjalta perusjoukon rajat määrittyvät tiukemmiksi: esimerkiksi peruspopulaationa on tupakoitsijat, ja joukkoa typistetään iällä ja muutamalla muulla muuttujalla. Näiden sääntöjen takia joukko pienenee selvästi, joten yleistettävyys samalla pienenee. Kompromisseja on kuitenkin tehtävä, koska muuten ei saada käytännössä toteuttamiskelpoista aineistopoimintaa. (Fink, 1995, 27-29)

Jos lähtöjoukko on epämääräisesti tai liian abstraktisti rajattu, on hankala eksplisiittisesti sanoa kaikkien alkioiden kohdalla, kuuluuko se itseasiassa joukkoon. Tämä liittyy ongelmaan perusjoukon koon määrittämisestä. Usein käyttökelpoiseksi tavaksi tuleekin perusjoukon kehyksen ottaminen jonkin valmiiksi kerätyn listan perusteella.

Kuten tässä tutkimuksessa on käynyt ilmi, aineiston hankinta saattaa tuottaa suuria ongelmia. Jos Suomen mittakaavassa tarkastellaan bioteknologia-alan toimijoita, esimerkiksi yritysten operatiivista johtoa, puhutaan melko pienestä toimijajoukosta. Kun joukko on pieni, voidaan lähteä liikkeelle tavoitteesta kerätä tiedot kaikista toimijoista. Jos tämä ei kuitenkaan onnistu, on selkeästi kuvattava, missä määrin saatu näyte on vinoutunut, ja arvioida tuloksia sen perusteella.

Toisaalta jos puhutaan suuremmasta joukosta ja tilastollisesti luotettavien tunnuslukujen laskemisesta, on ensin määriteltävä perusjoukko huolellisesti, ja tämän jäl-

keen on yritettävä tehdä mahdollisimman hyvä satunnaismoiminta.

3.3 Aineiston esikäsittely

Aineiston esikäsittelyssä suoritetaan lähdemateriaalille kvantifiointi, eli muokataan se sellaiseen muotoon, josta voidaan laskea erilaisia tunnuslukuja ja käsitellä sitä numeerisesti. Tämä esikäsittely on tehtävä osittain käsityönä. Luokittelua helpottamaan ja aineiston tallennusta varten laadittiin tietokonesovellus.

3.3.1 Tarvittavien muuttujien valinta

Aineiston kvantifiointi tarkoittaa tässä yhteydessä materiaalin luokittelua niin, että jokainen urapolku saadaan numeeriseen muotoon. Luokat on määriteltävä erikseen toimille, aloille sekä organisaatioille. Kukin aineistossa oleva henkilö saa yksikäsitteisen järjestysnumeron, joihin tarvittaessa voidaan viitata. Samoin menetellään organisaation, aseman ja koulutusluokkien kohdalla.

Aineiston perusluokitus tapahtuu perustietojen, koulutuksen sekä urapolkujen osalta. Organisaatiolla tarkoitetaan pienintä organisaatiota, joka on toimijaa lähinnä. Esimerkiksi yliopistossa tämä voisi käsittää laitoksen tai jopa niinkin pienen yksikön kuin tutkimusryhmän, mikäli on mielekästä ottaa tarkasteluun se suuremman yksikön asemesta.

Henkilön asemaa on hyvin vaikea määrittää yksikäsitteisesti. Esimerkiksi yliopiston professori voi olla samaan aikaan tutkija ja esimiesasemassa oleva henkilö. Muis-
sakin luokituksissa ilmenee jonkin verran epätarkkuuksia. Mikäli johonkin muuttu-
jaan ei saada arvoa tai sen tulkinta jää hyvin kyseenalaiseksi aineiston perusteella,
luokitellaan se omaan luokkaansa: "ei tiedossa". Joitakin analyysijä varten on ken-
ties järkevää korvata puuttuvat tiedot esimerkiksi datajoukon keskiarvolla. Tällöin
nämä eivät aiheuta vääristymiä asteikolla.

Ennenkuin esitellään tutkimuksessa käytettävät muuttujat, käsitellään muuttujatyyp-
pien luokittelua. Tässä tutkimuksessa käytetään muuttujatyyppejä, jotka ovat ylei-

sesti käytössä sosiaalitieteellisissä kvantitatiivisissa analyyseissä. Muuttujatyypit ovat nominaalinen, ordinaalinen, intervallinen ja suhdelukuasteikollinen muuttuja. Nominaaliasteikko, eli laatueroasteikko, kuvaa vain samanlaisuutta ja erilaisuutta, ei lainkaan esimerkiksi suuruuseroja. Nominaalista muuttujaa kutsutaan myös kategoriseksi. Esimerkki nominaalimuuttujasta on sukupuoli tai maantieteellinen alue. Myös tutkimuksessa käytettävä koulutusala on nominaalinen, koska sen arvoja ei voida asettaa suuruusjärjestykseen. Sen sijaan korkein suoritettu tutkinto ei ole enää nominaalinen, koska nämä muuttuja-arvot voidaan asettaa järjestykseen tutkinnon asteen mukaan. Tämä kuuluu seuraavaan ryhmään, joka on järjestysasteikolliset eli ordinaaliset muuttujat. Arvojen väliset etäisyydet eivät ole vielä määriteltäviä tässä ryhmässä, vaan ainoastaan arvojen keskinäinen järjestys on merkitsevä. Välimatka-asteikko eli intervalliasteikko on seuraavan muuttujatyypin skaala. Välimatka-asteikollinen muuttuja on esimerkiksi viidellä asteella ilmaistava "mielitymys johonkin": 1=hyvin vähäinen – 5=erittäin suuri, jossa välimatka 1-2 edustaa samaa erotusta kuin esim. välimatka 3-4. Tämän muuttujatyypin kategorioiden etäisyydet säilyvät samana, vaikka niitä käsiteltäisiin lineaarimuunnoksella, so. $y = ax + b$. (Sharma, 1996, 2-3). Viimeisenä tässä esitettävänä muuttujatyypinä on suhdelukuasteikollinen muuttuja, joka voi olla esimerkiksi ikä. Siinä on tarkkaan määrätty nollakohta, jota ei voi muuttaa. Sen takia ainoa sallittava operatio tälle muuttujalle on kiinteän nollakohdan säilyttävä muunnos: $y = ax$. (Alkula *et al.*, 1994, 85). Välimatka-asteikollisista tai suhdelukuasteikollisista muuttujista käytetään yhteisnimitystä metriset muuttujat.

Taulukkoon 3.2 on listattu aineistosta kvantifioitavat perusmuuttujat, jotka sisältävät syntymävuoden, sukupuolen ja koulutustiedot. Muuttujan tyyppi tarkoittaa äsken kuvattuja muuttujarajoituksia, puuttuva tieto tarkoittaa koodia, jolla ilmoitetaan numeerisessa datassa puuttuva alkio.

Mitkä muuttujat urapolkuaineistosta ekstraktoidaan, on arvioitava huolellisesti. Kaikia mahdollisia datasta ilmeneviä muuttujakandidaatteja ei kannata ottaa mukaan ilman arviointia. Kaikista urapoluista ei välttämättä saada kyseisen muuttujan informaatiota joko ollenkaan tai riittävän tarkasti. Jos kyseessä on kuitenkin tutkimuskäsitteiden operationalisoinnin kannalta merkittävä tieto, on mahdollista kier-

tää tätä hieman asettamalla puuttuvat arvot esimerkiksi koko aineiston keskiarvoksi. Jos toisaalta aineistossa esiintyy informaatiota, jota ei heti tarvita, mutta jonka voidaan ajatella osoittautuvan tutkimuksen myöhemmässä vaiheessa tärkeiksi tai joi- ta voidaan hyödyntää jatkotutkimuksissa, ne kannattaa kvantifioida samalla. Tässä tutkimuksessa luokittelusovellus tallentaa tiedot tietokantaan, jota on helppo myö- hemmin hyödyntää.

Esimerkiksi jos tutkimusjoukkona olisivat pelkästään tutkijat, mielekkäitä koulutus- ja uratiedoista saatavia muuttujia voisivat olla esimerkiksi tutkinnon kesto ja tutki- jan julkaisujen määrä, mikäli nämä ilmenevät aineistosta.

Muuttujan nimi	Muuttujan tyyppi	Puuttuva tieto
Syntymävuosi	suhdeluku	0
Sukupuoli	nominaalinen	0
Korkein suoritettu tutkinto	ordinaalinen	0
Koulutusalan luokitus	nominaalinen	0
Koulutusala täsmällisesti	nominaalinen	0
Tutkinnon suoritusvuosi	suhdeluku	0

Taulukko 3.2: Henkilön perustiedot P_i

Taulukossa 3.2 on siis listattuna henkilön perustiedot ja koulutus P_i , missä i on henkilön indeksi. Ura-askeleen kuvaamiseen käytetään toista tietorakennetta A_j , missä j on ura-askeleen indeksi tietyllä henkilöllä.

Taulukossa 3.3 on muuttujat, joita tarvitaan yhden ura-askeleen kuvaamiseen. Yk- sittäiset ura-askeleet muodostavat vektorimuuttujan kunkin henkilön kohdalla. Täs- tä vektorista voidaan laskea helposti esimerkiksi kokonaisliikkuvuus (eli työposi- tioiden määrä) tai muita tarvittavia tietoja.

Organisaatityyppi luokittelee ylemmällä tasolla organisaation (esim. yritys tai kor- keakoulu). Täsmällinen organisaatio antaa tarkemman rajauksen, eli esimerkiksi *Firma Oy* tai *Helsingin yliopisto*. Tämän muuttujan frekvenssien avulla etsitään analyysissä keskeisiä organisaatioita, joiden kautta useat toimijat ovat kulkeneet uransa aikana. Yksikön toimiala luokittelee jokaisen organisaatioyksikön toimia- lan.

Otetaan esimerkiksi suuri julkisen sektorin tutkimuslaitos, jonka eräässä yksikössä henkilö on töissä. Yksikkö tekee farmasian alan tutkimusta. Henkilö toimii siellä tutkimuspäällikkönä eli käytännössä johtaa tutkimusprojekteja ja toimii hallinnollisena päällikkönä. Organisaatiotyyppi on tällöin julkinen tutkimuslaitos, ja yksikön toimialaksi tulee farmasia, vaikka pääorganisaatio on monialainen.

Henkilön toimiasema on hieman ongelmallinen. Tässä tapauksessa henkilö on sekä tutkija että esimies. Kuitenkin esimiesasema on tässä merkitsevä, joten toimeksi määräytyy esimies. Yhden käsitteellisen ongelman muodostaa se, että suuressa organisaatiossa esimiehen asema saattaa olla paljon vastuullisempi kuin pienen organisaation jonkin operatiivisen johtajan asema. Tätä voitaisiin ilmentää ottamalla mukaan organisaation kokoa määrittävä muuttuja. Tätä tietoa ei kuitenkaan ole materialissa saatavissa, vaikkakin monien organisaatioiden koko voitaisiin määrittää jälkikäteen. Kuitenkin koska puuttuvia tietoja jäisi paljon ja erillinen selvitys olisi melko työläs, ei tähän ole ryhdytty.

Henkilön toimiala kuvaa hänen toimensa, eikä koko organisaation toimialaa. Toisin sanoen onko kyseessä talous- tai hallinnolliset tehtävät, tuotekehitys vai jonkin alueen tutkimus. Tämä voi tietenkin poiketa organisaation päätoimialasta.

Ulkomaa -muuttuja kertoo yksinkertaisesti, onko toimi ulkomailla, eli tämän pohjalta voidaan laskea henkilön ulkomaankokemusta päätoimien osalta.

Muuttujan nimi	Muuttujan tyyppi	Puuttuva tieto
Organisaation luokka	nominaalinen	0
Täsmällinen organisaatio	nominaalinen	0
Yksikön toimiala	nominaalinen	0
Henkilön toimen luokka	nominaalinen	0
Henkilön toimi	nominaalinen	0
Toimiala	nominaalinen	0
Ulkomaan toimi	nominaalinen	0
Toimen alkamisaika	suhdeluku	0
Toimen kesto vuosissa	suhdeluku	0

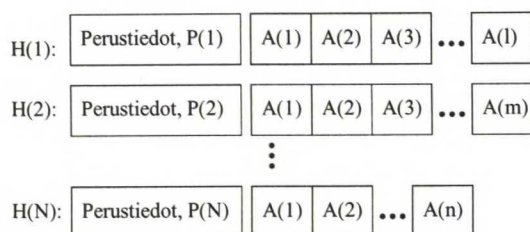
Taulukko 3.3: Yksittäisen ura-askeleen tiedot A_j

Jotta datan hahmottaminen selkenee, kaavassa 3.1 esitetään tietorakenne yhden henkilön (H_i) perustiedoista ja ura-askelista:

$$H_i = \{P, \{A_1 A_2 \dots A_n\}\} \quad (3.1)$$

Kyseessä olevan henkilön urapolun pituus on n . Perustiedot on määritelty rakenteessa P ja yksittäiset ura-askleet muodostavat joukon $A_j, j = 1 \dots n$

Koko kvantifioitu urapolkuaineisto kaikista henkilöistä (koko N) on puolestaan vektori, jonka elementit ovat kyseisiä henkilötietorakenteita: $H_1, H_2 \dots H_N$. Kuvassa 3.2 on visualisoitu tilannetta.



Kuva 3.2: Visualisointi tietorakenteista

Kuvan 3.3 listauksessa on esimerkki fiktiivisen henkilön urapolkudatasta kvantifioituna ja muunnettuna Matlabin ymmärtämään tietorakennemuotoon. Kyseessä on yksi Matlab-kielinen tietorakenne, jotka asetetaan analyysissä *data* -nimiseen taulukkoon. Tietorakenteessa näkyvät viisi ensimmäistä muuttujaa ovat henkilötiedot ja koulutustiedot. Seuraavat kuusi vektorimuotoista muuttujaa ilmentävät urapolkudataa askel askeleelta.

Tästä kvantifioidusta aineistosta saadaan edelleen laskettua tarvittavia lisämuuttujia analyysin tarpeisiin. Esimerkiksi organisaatioliikkuvuus on kaikkien henkilön urapolulla esiintyvien organisaatioiden kokonaismäärä. Ulkomaankokemusta voidaan vastaavasti mitata sen mukaan, kuinka monta vuotta henkilö on eri toimien puitteissa ollut ulkomailla. Seuraavassa alaluvussa on kuvattu kunkin muuttujan luokitus, eli kuinka todellisesta aineistosta on päästy näiden muuttujien rajaamaan kvantitatiiviseen aineistoon.


```
data(1) = struct('person_id', 1, ...
                'birth_year', 1962, ...
                'education', 1, ...
                'profession', 1, ...
                'edu_year', 1990, ...
                ...
                'position_ids', [ 1 2 3 4 5 ], ...
                'positions',    [ 8 1 2 2 2 ], ...
                'professions',  [ 3 2 2 2 3 ], ...
                'organizations', [ 1 1 1 2 2 ], ...
                'begin_years',  [ 1986 1988 1989 1989 1993], ...
                'durations',    [ 1.3 5 0.5 1 3 ] );
```

Kuva 3.3: Yksittäisen ura-askeleen tietorakenne A_i

3.3.2 Aineiston kvantifiointi muuttujien arvoiksi

Tässä tutkimuksessa käytettävän aineiston määrä on siinä määrin pieni, että analyysi on mahdollista (joskin työlästä) tehdä käsinkin. Tarkoituksena on kuitenkin osoittaa, kuinka analyysi saadaan tehokkaasti tehtyä tietokoneohjelmiston avulla, jolloin analyysikoneistoa on helppo muokata ja käsitellä monenlaista urapolkuihin liittyvää dataa. Aineiston esikäsittely aiheuttaa sen, että kovin suuret datamäärät tekevät käsittelystä työlään. Toinen tapa kerätä aineistoa olisi valmis kyselylomake, johon tutkimukseen osallistuvat voivat itse luokitella aineiston. Tällöin on laadittava kuitenkin ennakolta valmis rajausta, jonka käyttökelpoisuus nähdään vasta, kun aineisto on lopulta käsillä. Jos taas aineisto on luokittelematon, voidaan sopivia luokituksia kokeilla. Varsinkin uudentyyppisen aineiston ja menetelmän ollessa kyseessä luokituksen kokeilu on väistämätöntä.

Numeromuotoisen aineiston muokkaaminen käsin on virhealtista ja hidasta, joten työkalun kehittäminen tähän tarkoitukseen oli perusteltua. Aineiston luokittelun helpottamiseksi laadittiin tietokonesovellus, jonka avulla informaation syöttäminen on nopeampaa ja vähemmän virhealtista kuin numeeristen luokkien käsin kirjaaminen. Luokittelu- ja tietojen tallennussovelluksen avulla saadaan aineisto helposti kvantifioitua ja tallennettua sellaiseen muotoon, että se voidaan lukea suoraan analyysiohjelmistoon. Ideana on, että sovellukseen voidaan syöttää kunkin henkilön

tiedot niin, että luokat valitaan valmiista listasta tai tarvittaessa lisätään uusi luokka. Sovelluksen käyttäjä näkee luokat sanallisessa muodossa, ja ohjelma pitää itse kirjaa luokkien koodeista. Tällöin virheet, joita syntyisi numeeristen luokkien käsin syöttämisestä, poistuvat. Kun urapolut on syötetty, saadaan tiedot tulostettua sopivassa muodossa analyysivaihetta silmälläpitäen.

Sovellus on www-pohjainen ja käyttää MySQL-tietokantaa tietojen tallentamiseen. Ohjelma on kirjoitettu PHP-kielellä ja se toimii www-palvelimella. Sovellusta on mahdollista kehittää niin, että useat eri tietokoneilla työskentelevät tutkijat voivat syöttää aineistoa samaan tietokantaan helposti www-selaimen avulla. Tiedot ovat keskitetyksi yhdellä palvelimella. Tämän tutkimuksen puitteissa tehty sovellus on kuitenkin vielä melko alkeellinen, mutta sitä voidaan kehittää eteenpäin. Tarkempi esittely luokittelusovelluksen käyttöliittymästä on liitteessä B.

Aineistossa oli joitakin puuttuvia tietoja, tai sitten saadusta tiedosta ei voitu riittävällä tarkkuudella selvittää järkevää muuttujan arvoa. Kvantifioinnissa tällaiset muuttujan instanssit merkittiin yleensä nollalla, missä se oli järkevää (ks. aiemmat muuttujataulukot). Tällä ratkaisulla analyysissä tunnuslukuja laskettaessa huomataan puuttuvat tiedot ja voidaan tarvittaessa korvata puuttuva tieto esimerkiksi muuttujan keskiarvolla, jos se on tarpeellista.

Aluksi aineisto kirjoitettiin puhtaaksi ja tarkastettiin, ovatko jokaisen aineistoalkion tiedot riittävät. Jos puutteita oli liikaa, tällaiset alkiot poistettiin aineistosta. Kustakin ansioluettelosta otettiin perustiedot: syntymävuosi ja koulutustiedot sekä kirjattiin yksikäsitteinen urapolku aiemmin kuvattujen muuttujien suhteen. Päällekkäiset toimet kirjattiin myös silloin, kun ei voitu erotella selkeästi kumpi on päätoimi ja kumpi sivutoimi. Alle puolen vuoden toimet poistettiin aineistosta niiden oletetun marginaalisen vaikutuksen vuoksi.

Näiden vaiheiden jälkeen luokiteltiin organisaatiot, alat sekä toimet. Organisaatiot luokiteltiin hyvin väljästi – tarve lisämuuttujille, joissa luokitus on tarkempi, on perusteltu. Sitä ei kuitenkaan vielä tässä vaiheessa tehty.

Toimien luokittelussa on useita ongelmia. Ensinnäkin eri organisaatioissa sama toiminimike saattaa sisältää hieman erilaisen tehtävätyypin ja vastuualueen. Toisaalta täysin sopivia luokituksia ei kaikkiiin toimiin ole löydettävissä, jos luokkajaosta ha-

lutaan riittävän väljä. Kun aineisto on luokiteltu ja kvantifioitu, sitä voidaan käyttää analyysiohjelmiston syötteenä.

3.4 Analyysiohjelmisto

Analyysi suoritetaan Matlab-matematiikkaohjelmiston avulla. Matlab sisältää laajat kirjastot tilastolliseen laskentaan liittyviä funktioita sekä monipuolisen ohjelmointikielen.

Matlab ei ole ainoa mahdollisuus analyysin tekniselle toteutukselle, mutta monipuolisuutensa sekä hyvän ohjelmoitavuuden takia järkevä vaihtoehto. Verkostoanalyysien toteuttamiseen on olemassa valmiita tietokoneohjelmistoja, mutta valmiin analyysipaketin käyttäminen saattaa aiheuttaa tiettyjä rajoitteita tutkimukselle. Voi osoittautua, että ohjelmisto on rajoittunut vain tiettyntyyppisiin aineiston käsittelyihin, jolloin mahdollinen lisäanalyysi olisi joka tapauksessa tehtävä jollakin muulla ohjelmalla. Matlabiin on ohjelmoitavissa varsin suoraviivaisesti monenlaisia analyysiproseduureja. Tutkijan taidoista riippuen mahdollisuudet erilaisten analyysisovellusten tekemiselle ovat erittäin laajat. Aineiston ja tulosten visualisointi on myöskin laajojen funktiokirjastojen johdosta Matlabilla varsin nopeasti toteutettavissa.

Taulussa 3.4 on esitetty analyysiproseduurit ja data jaettuna Matlab-tiedostoihin. Liitteessä A on listattu analyysivaiheiden Matlab-kieliset lähdekoodit.

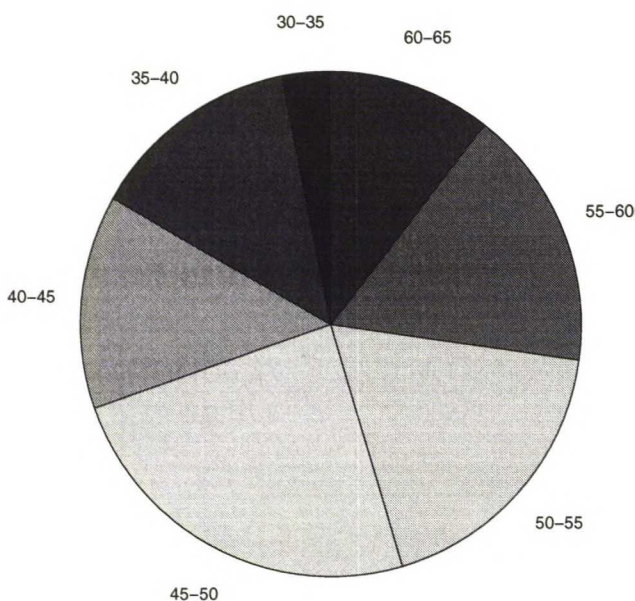
Tiedosto	Tarkoitus	Tyyppi
data_66.m	Kvantifioitu henkilödata	Data
class_names.m	Kaikkien luokitusten nimet	Data
run.m	Analyysin päätaso	Ohjelma
initialize.m	Tietojen luku ja analyysin alustukset	Ohjelma
compute_basic.m	Perustietojen laskenta	Ohjelma
compute_experiences.m	Kokemusmuuttujien laskenta	Ohjelma
pca.m	Pääkomponenttianalyysin laskenta	Ohjelma
cluster_pca.m	Pääkomponenttianalyysin jälkeinen klusterointi	Ohjelma
compute_networks.m	Verkostoanalyysin laskenta	Ohjelma
show*.m	Tietojen visualisointi ja esitys	Ohjelma

Taulukko 3.4: Analyysin Matlab-tiedostot

3.5 Perustulokset kohdejoukosta

Myöhempiä tarkasteluja ja tulkintoja varten on tässä vaiheessa järkevää esittää kohdejoukosta muutamia perustietoja.

Kohderyhmän iän keskiarvo on 49 vuotta, ja iät vaihtelivat välillä 30-64 vuotta. Jakaumakuvasta 3.4 nähdään, että kolme eniten edustettua ikäryhmää olivat välillä 45-60 vuotta, mikä on luonnollista, kun on kyse esimiesasemassa ja vaativissa asiantuntijatehtävissä olevista henkilöistä. Tällaiseen asemaan pääsy edellyttää kokemusta.

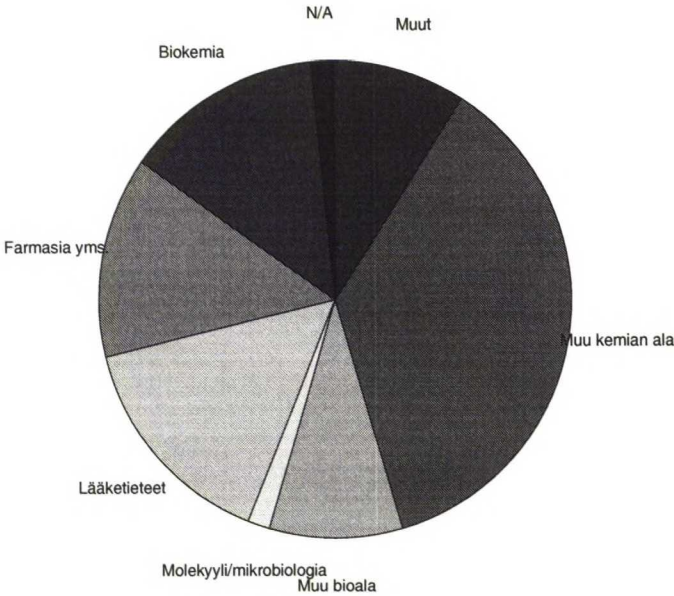


Kuva 3.4: Kohdejoukon ikäjakauma

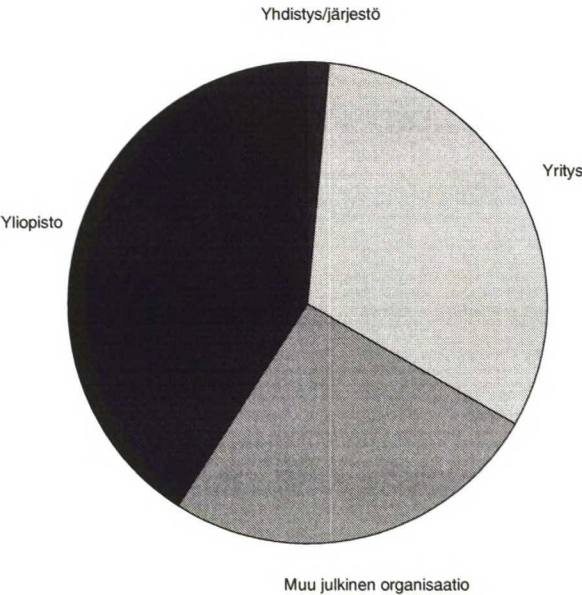
Koulutustason osalta valtaosalla henkilöistä oli tohtorin tutkinto, ja vain kymmenellä prosentilla oli korkeimpana tutkintona maisterin tutkinto. Koulutusalan suhteen suurimmalla osalla oli jokin kemian alan erikoistumiskohde, kuten orgaaninen kemia. Farmasia ja lääketieteet olivat myös melko hyvin edustettuina, kuten kuvajasta 3.5 nähdään.

Kohdehenkilöiden tehtävä (viimeisen työposition mukaan) oli pääasiassa johtotehtävät (67 prosenttia) ja 32 % toimi asiantuntijatehtävissä. Kuvaajasta 3.6 nähdään, että yliopiston tai korkeakoulun palveluksessa oli 27 % vastanneista, 17 % muussa

julkisessa organisaatiossa (valtio, kunta tms), sekä 21 % yrityksissä.



Kuva 3.5: Kohdejoukon koulutusjakauma



Kuva 3.6: Kohdejoukon viimeisen työpaikan jakauma

3.6 Pääkomponenttianalyysi uratyypin hahmottamisessa

Aineistossa esiintyvien uratyypin ja urien yhtenäisten ominaisuuksien arvioinnissa käytetään pääkomponenttianalyysia (Principal Component Analysis, PCA). Menetelmä mahdollistaa suuren havaintoaineiston redusoinnin niin, että suuri joukko havaintojen muuttujia voidaan ilmaista pienemmällä muuttujajoukolla, jotka kuitenkin ilmaisevat olennaisen informaation koko aineistosta. Tätä uutta muuttujajoukkoa kutsutaan alkuperäisen aineiston pääkomponenteiksi. Pääkomponenttianalyysi on monimuuttujamenetelmä, jossa muuttujajoukko voi riippua toisesta.

Esimerkki pääkomponenttianalyysistä on yritysanalyysi, jossa selvitetään yritysten kuntoisuuksia. Jos lähtökohtana on 100 erilaista taloudellista mittaria, menetelmän avulla lasketaan vaikkapa kaksi komposiitti-indeksiä, joiden avulla on helpompi tehdä tarkasteluja ja arviointeja. Nämä indeksit ovat painotettuja summia alkuperäisistä mittareista, ja käsittävät olennaisen osan lähtömuuttujajoukon informaatiosta. Olennainen määräytyy yleisesti tarkastelunäkökulman ja ongelman laadun mukaisesti. Alkuperäinen ongelma siis redusoituu näiden painokertoimien etsimiseen. (Sharma, 1996, 58)

3.6.1 Menetelmän teoreettinen tarkastelu

Analyttisesti pääkomponenttimenetelmä voidaan kuvata seuraavalla tavalla:

$$\begin{aligned}\xi_1 &= w_{11}x_1 + w_{12}x_2 + \dots + w_{1p}x_p \\ \xi_2 &= w_{21}x_1 + w_{22}x_2 + \dots + w_{2p}x_p \\ &\vdots \\ \xi_p &= w_{p1}x_1 + w_{p2}x_2 + \dots + w_{pp}x_p\end{aligned}\tag{3.2}$$

$$\begin{aligned} w_{i1}^2 + w_{i2}^2 + \dots + w_{ip}^2 &= 1, & i &= 1, \dots, p \\ w_{i1}w_{j1} + w_{i2}w_{j2} + \dots + w_{ip}w_{jp} &= 0, & \text{kaikille } i &\neq j \end{aligned} \quad (3.3)$$

Yhtälöissä 3.2 merkintä ξ_i on i :s pääkomponenttivektori, joita tulee yhteensä p kappaletta eli sama määrä kuin alkuperäisiä muuttujia. Alkuperäisiä muuttujia ovat x ja muuttujan j painokerroin pääkomponentille i on w_{ij} . Matemaattisesti ongelmana on valita painokertoimet niin, että ensimmäinen pääkomponentti saa suurimman osuuden alkuperäisen datan varianssista, ja toinen pääkomponentti saa suurimman osuuden jäljellä olevasta varianssista. Jos tarkastellaan kaikkia pääkomponentteja yhdessä, ne kattavat koko aineiston varianssin. (Sharma, 1996, 66)

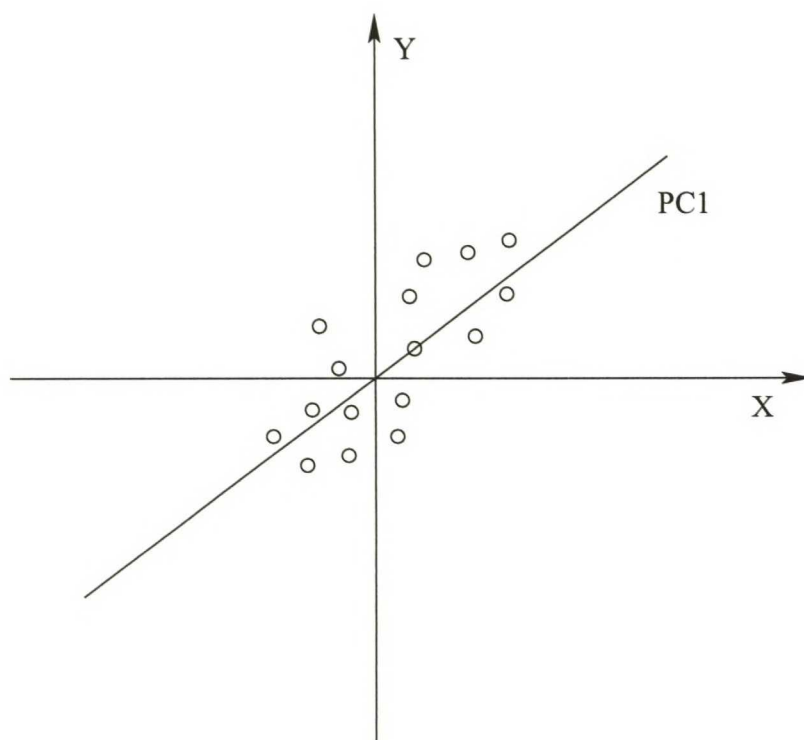
Yhtälöt 3.3 määräävät rajoitteet pääkomponenttien muodostukselle. Ensimmäinen yhtälö rajoittaa painokertoimia (w) kasvamasta liian suuriksi. Toinen yhtälö varmistaa, että komponenttiakselit ovat kohtisuorassa toisiaan vastaan.

Ideana PCA-analyysissä on vähentää lähtöjoukon muuttujadimensiota. Usein monimuuttujadatan kohdalla on niin, että muuttujilla on yhteisvaihtelua ja tämän yhteisvaihtelun selittää vain muutama tekijä, jotka voidaan saada esille sopivalla muunnoksella. Pääkomponenteista valitaan tietty osa (maksimissaan niitä on sama määrä kuin alkuperäisiä muuttujia) ja vähennetään tällä tavalla alkuperäisen datan dimensiota.

Kuvassa 3.7 on yksinkertainen kaksiulotteinen esimerkki. Ideana on muuntaa kaksiulotteinen data yksiulotteikseksi niin, että kuvassa oleva suora PC1, eli 1. pääkomponentti, saisi suurimman varianssin datasta. Kuvassa olevat pisteet projisoidaan suoralle. On huomattava, että kun pisteet projisoituvat toistensa päälle, niitä ei voida enää erottaa pääkomponenttiesityksessä toisistaan. Tämän takia tapahtuu informaatiohäviötä, joka olisi hyvä ottaa huomioon ja yrittää arvioida sen suuruus sekä vaikutukset, kun tehdään johtopäätöksiä tuloksista.

PCA-menetelmää varten ei voida käyttää suoraan kvantifioituja urapolkutietoja, vaan niitä on käsiteltävä lisää. Muodostetaan niin sanottu havaintomatriisi, jossa konstruoidut muuttujat ovat toisella akselilla ja havaintojen numerot toisella. Urapolut ovat eri pituisia, joten urapolkutiedoista on laskettava sopivat tunnusluvut, eli

ilmaistava haluttu data havaintomatriisiin sopivaksi. Tätä kuvataan tarkemmin seuraavassa alaluvussa.



Kuva 3.7: Kaksiulotteinen esimerkki

3.6.2 Tarvittavien muuttujien konstruointi

Koska pääkomponenttimenetelmää varten tarvitaan sama määrä muuttujiahavaintoja jokaisesta alkioista (henkilöstä), kehitetään jokaisesta urapolusta joukko muuttujia, jotka kuvaavat uraa. Tällöin saadaan muodostettua havaintomatriisi A_{ij} , jossa henkilöt $i = 1..N$ muodostavat toisen indeksin ja näihin liittyvät muuttujat $j = 1..M$ toisen.

Käytetään seuraavanlaisia muuttujia, jotka lasketaan urapolkudatasta:

- 1. Ulkomaankokemus vuosissa.
- 2. Akateeminen työkokemus (yliopistoura) vuosissa
- 3. Kokemus yritysmaailmassa vuosissa
- 4. Tutkimuskokemus vuosissa
- 5. Johtamiskokemus vuosissa

Merkitään edellä määritettyjä muuttujia järjestyksessä x_1-x_5 . Aineistosta voidaan nyt muodostaa matriisi, jota on havainnollistettu taulukossa 3.5 (muuttujien arvoja ei ole tässä merkitty näkyviin).

	x_1	x_2	x_3	x_4	x_5
Henkilö 1					
Henkilö 2					
Henkilö 3					
\vdots					
Henkilö N					

Taulukko 3.5: Havaintomatriisi

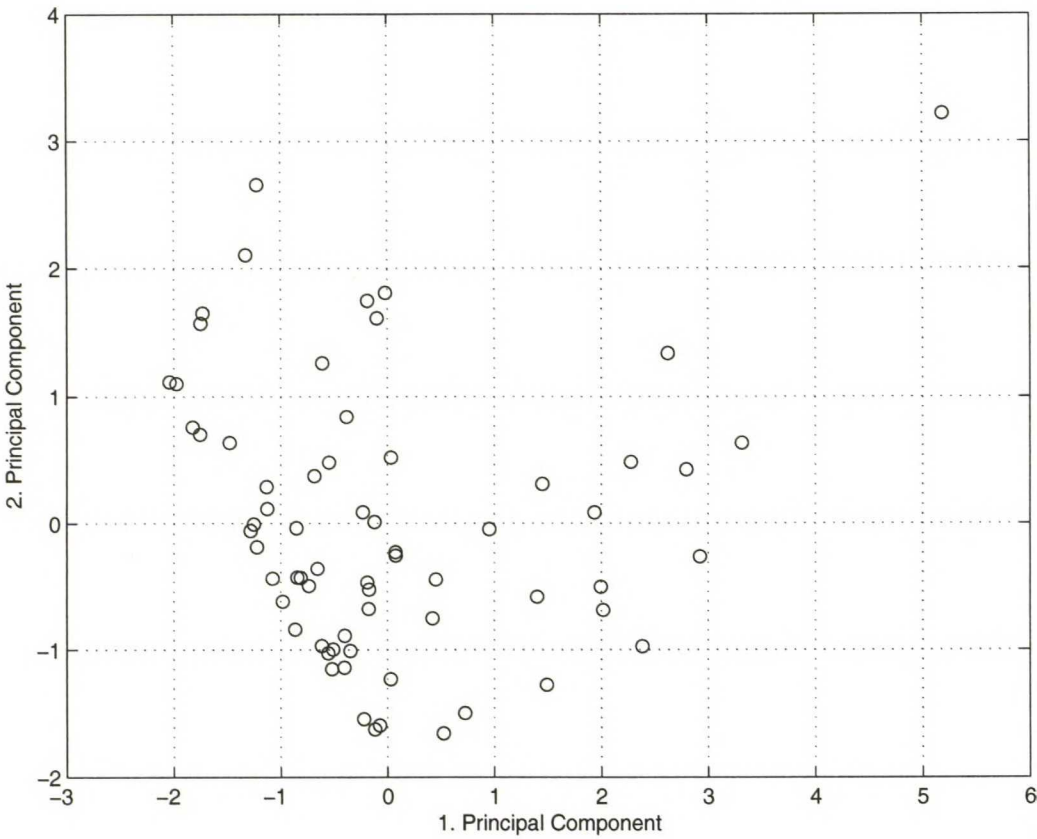
PCA-analyysillä tehdyn redusoinnin jälkeen viisiulotteinen data saadaan kaksiulotteiseksi. Taulukossa 3.6 pääkomponentit PC_1 ja PC_2 sisältävät suurimman varianssin aineistosta.

	PC_1	PC_2
Henkilö 1		
Henkilö 2		
Henkilö 3		
\vdots		
Henkilö N		

Taulukko 3.6: Redusoitu havaintomatriisi

3.6.3 PCA-analyysin toteutus

Kuvassa 3.8 on esitetty PCA-analyysin tulokset piirtämällä saadut kaksiulotteiset pisteet koordinaatistoon.



Kuva 3.8: PCA-analyysin tulokset

Pääkomponenttien painokertoimet antavat viitettä siitä, että ensimmäinen pääkomponentti ilmaisee vastaavuutta muuttujiin 3 ja 5, eli kokemus yritysmaailmassa sekä johtamiskokemus. Toinen pääkomponentti näyttäisi ilmaisevan muuttujia 1, 2 ja 4, eli kansainvälinen kokemus, akateeminen ja tutkimuskokemus. On huomattava, että aina pääkomponenttien painokertoimien antamat viitteet vastaavuuksista alkuperäisiin muuttujiin eivät ole selkeästi havaittavissa.

Koordinaatistoakseleiden keskelle sijoittuva piste tarkoittaa sellaista henkilöä, jolla nämä kaksi pääkomponenttimuuttujaa ovat keskimääräisiä koko aineistoa ajatellen. Suuret positiiviset arvot ensimmäisellä pääkomponentilla (x-akselilla) tarkoittavat merkittävää kokemusta yritysmaailmasta ja johtamisesta ja kääntäen negatiiviset arvot pientä vastaavaa kokemusta.

Kun havainnot on saatu PCA:n avulla kahteen ulottuvuuteen, on mielekästä selvittää minkälaisia ryhmiä eli klustereita nämä muodostavat. Näin pienellä aineistolla klustereiden yksikäsitteinen selvittäminen ei ole tarkkaa, mutta suuremmalla aineistolla saataisiin todennäköisesti luotettavampia tuloksia, toisin sanoen pisteiden määrän kasvaessa on todennäköisempää, että klusterit hahmottuvat paremmin kohinan seasta. On tietysti mahdollista, että klustereita ei muodostu lainkaan, joka on sekin uutta informaatiota.

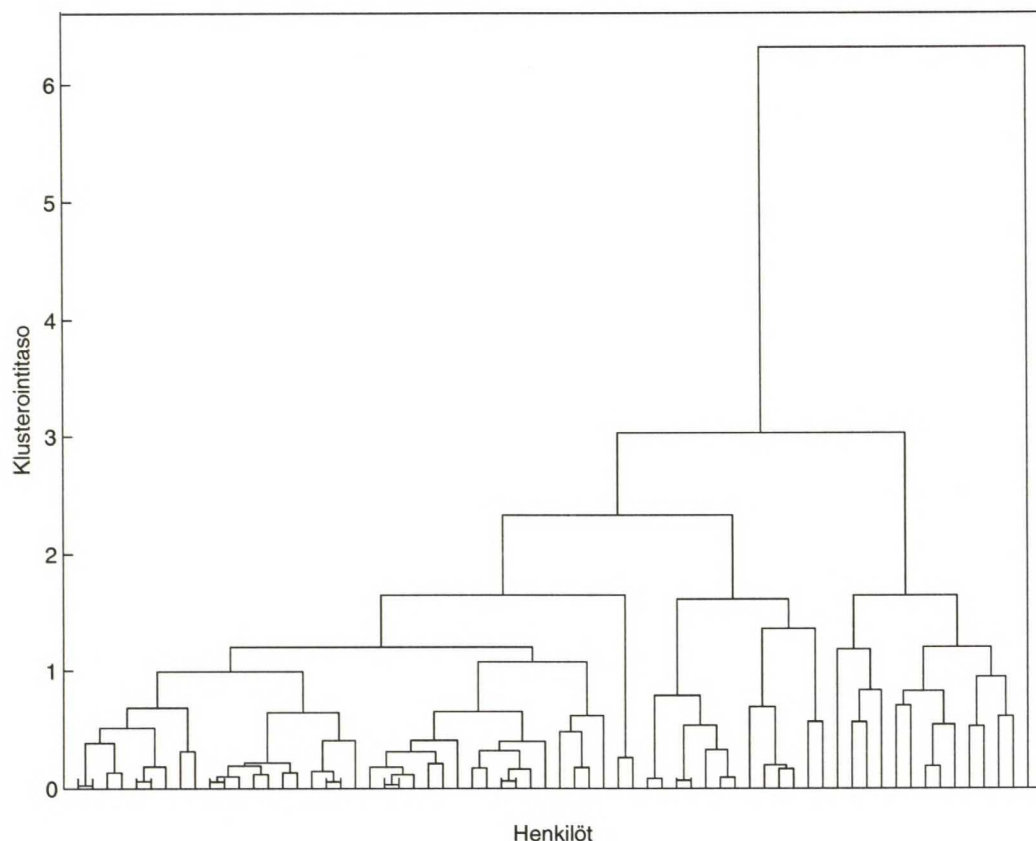
Kaksiulotteisen PCA-analyysin tulosten klusteroinnissa käytetään yksinkertaista pisteiden läheisyyteen perustuvaa menetelmää. Jokaisen pisteen välinen etäisyys mitataan, ja näistä etäisyyksistä muodostetaan puumainen esitys. Tätä hierarkkista esitystä nimitetään dendrogrammiksi.

Käsittely etenee vaiheittain. Ensin lasketaan jokaisen pisteen välinen etäisyys toisista käyttäen normaalia euklidista etäisyysmittaa, joka on formuloitu kaavassa 3.4.

$$d_{jk} = \sqrt{\sum_{i=1}^N (x_{ji} - y_{ki})^2}. \quad (3.4)$$

Tämän jälkeen muodostetaan hierarkkinen klusterointi saaduista etäisyyksistä. Lähekkäiset pisteet tai objektit yhdistetään binääriklustereiksi, ja näin saatuja saatuja objekteja edelleen yhdistetään ylemmällä tasolla ryhmiin. Lopulta kaikki pisteet,

pisteparit ja korkeamman tason objektit on yhdistetty hierarkkiseksi puuksi. Jotta edellä kuvattu menetelmä selkenisi, esitetään kuvassa 3.9 dendrogrammi laskettuna PCA-analyysin tuloksista.



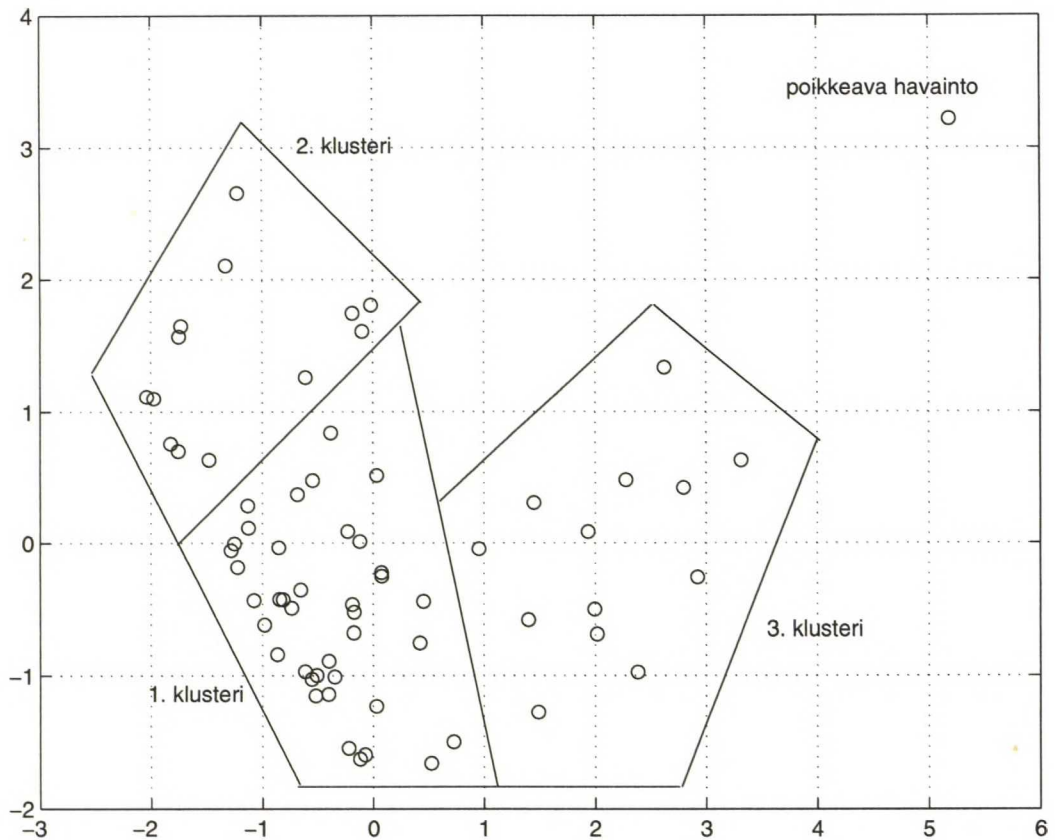
Kuva 3.9: Dendrogrammi $N = 66$

Kuten kuvasta 3.9 voidaan nähdä, poikkeava havainto (outlier) on kaikkein kauimpana muista pisteryhmistä, joten siihen liittyvä binääripuun haarautuminen tapahtuu aivan puun juuressa (eli kuvan yläreunassa). Kaikkein matalimmat käännteisen U:n muotoiset parit tarkoittavat hyvin lähellä toisiaan olevia yksittäisiä pisteitä. Algoritmi on lähtenyt liikkeelle näistä pareista ja yhdistellyt tämän jälkeen suurempia objekteja toisiinsa, kunnes on päästy puun juureen. Tultaessa kuvaa alaspäin puun juuresta nähdään kohta, jossa haaraumat muodostavat neljä vertikaalista janaa. Tämä on se kohta, jolta on otettu analyysissä havaitut klusterit: yksi suurempi ryhmä vasemmalla, sitten kaksi pienempää ja lopuksi outlier.

Kuvassa 3.10 nähdään tulokset. Tuloksista on pystytty muodostamaan kolme klusteria sekä kuvaajan oikeassa yläreunassa näkyvä poikkeava arvo, jota ei täysin muista poikkeavien arvojensa takia kannata ottaa mukaan tulosten myöhemmässä arvioinnissa. Toisaalta jos outlier-havaintoja on useita, niitä voidaan ottaa jossakin tapauksessa lähempään tarkasteluun, koska ne saattavat sisältää kiinnostavaa informaatiota.

Suurin klusteri näyttää olevan vasempaan alareunaan sijoittuva klusteri, joka sisältää 39 havaintopistettä. Kaksi muuta ryhmää ovat huomattavasti pienempiä ja sisältävät kumpikin 13 havaintopistettä. Taulukossa 3.7 taulukoitu saadut arvot.

Taulukkoon on nimetty kolme pääklusteria, ja niille on laskettu alkuperäisten viiden kokemusmittareiden pohjalta keskiarvot. Ensimmäinen klusteri on nimetty "tyypilliseksi", koska se on näistä kolmesta selvästi suurin. Siitä voidaan panna merkille,



Kuva 3.10: PCA-analyysin tulokset, klusterit

Nimi	1. Klusteri "Tyypillinen"	2. Klusteri "Akateeminen"	3. klusteri "Yrityskenttä"	"poikkeava"
1 Ulkomaan	0,74	2,58	0,77	20
2 Akateeminen	12,51	42,38	9,19	0
3 Yritys	2,01	1,04	20,46	33
4 Tutkimus	9,53	17,09	4,62	0
5 Johtamis	2,45	5,31	15,73	25

Taulukko 3.7: PCA-analyysin tulokset

että ulkomaankokemus ja johtamiskokemus ovat kaikkein pienimmät näistä ryhmistä. Oli jo alussa selvää, että akateeminen ja yrityspuoli tulisivat selkeästi esille analyysissä, koska ne olivat lähtödatassa nähtävänä erillisinä joukkoina.

3.7 Verkostanalyysi urapoluista

3.7.1 Lähtökohdat

Verkoston käsite ei ole kovin selkeästi määriteltävissä. Sitä voidaan käyttää yleisenä metaforana, kun puhutaan monimutkaisista, monien toimijoiden sosiaalisista järjestelmistä. Esimerkiksi voidaan puhua verkostotaloudesta. Yleiskielimäinen käyttö voi kuitenkin hämärtäessään käsitteitä vähentää uskottavuutta kaikkeen verkostotutkimukseen. Toinen tapa verkostojen käyttöön tutkimuksessa on pohjustaa jokin tieteellinen teoria verkostojen yhteyteen. Kolmanneksi verkosto-näkökulmaa voidaan soveltaa analyttisenä kvantitatiivisena menetelmäjoukkona, jolloin verkosto toimii lähestymistapana aineiston ja tutkittavan kysymyksen käsittelyssä (Johanson *et al.*, 1995, 4-5). Yleinen ja analyttisesti hedelmällinen tapa määritellä verkostanalyysi on nähdä se havaintoyksikköjen välisten suhteiden analyysinä (ks. esim. Johanson *et al.* (1995), Scott (2000), Wasserman ja Faust (1999)).

Tässä työssä verkostanalyysi nähdään kvantitatiivisena menetelmänä hahmottaa havaintoaineistoa. Erotellaan toimijat ominaisuuksineen sekä yhteydet toimijoiden välillä.

Tehtävänä on kartoittaa ja laskea henkilöiden organisaatioliikkuvuudesta tunnuslukuja. Tavoitteena on selvittää sellaisia organisaatioita, joissa on ollut useilla henkilöillä paljon samaan aikaikkunaan sijoittuvia toimia. Lisäksi kartoitetaan merkittäviä toimijoita.

3.7.2 Verkostanalyysin käsitteitä

Tämä verkostanalyysi pohjautuu *graafien* käsittelyyn. Graafi on yksinkertaistetusti joukko pisteitä ja niiden välisiä yhteysviivoja. Tämän luvun verkostanalyysissä tietyt henkilöiden väliset yhteydet kuvataan graafina.

Graafin pisteitä kutsutaan tässä solmuiksi, koska piste mieltyy herkästi matemaattiseksi pisteeksi, jonka ominaisuuksia ovat pelkät koordinaatit. Erotuksena matemaattisesta pisteestä, solmulla ei ole kiinnitettyjä koordinaatteja. Graafin solmuja

voidaan yhdistää pareiksi, joita kutsutaan särmiksi. Särmiä on yksisuuntainen yhteys solmusta v_1 solmuun v_2 . Yhteydet voivat olla myös kaksisuuntaisia, eli (v_1, v_2) ja (v_2, v_1) . Jos kaikki graafin yhteydet ovat kaksisuuntaisia, puhutaan *suuntaamattomasta graafista*, erotuksena suunnatusta graafista. (Weiss, 1997, 283-286)

Formaalisti graafi G voidaan kirjoittaa pariksi $G = (V, E)$, jossa on n kappaletta solmuja V ja joukko solmujen välisiä särmiä E :

$$V = \{v_1, v_2 \dots v_n\} \quad (3.5)$$

$$E = \{(v_1, v_2), (v_3, v_4), \dots, (v_l, v_k)\} \quad (3.6)$$

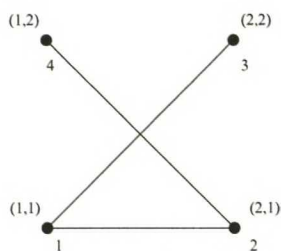
3.7.3 Analyysin toteutus

Seuraavassa selostetaan verkostanalyysin toteutus laskennallisesti. Tarkoituksena on selvittää toimijoiden välisiä yhteyksiä. Yhteys tarkoittaa tässä kahden toimijan samanaikaista toimimista samassa paikassa. Seuraavassa laskennallisessa analyysissä nämä kontaktit tai linkit tutkitaan sekä henkilöiden että organisaatioiden näkökulmasta.

Aluksi selvitetään henkilöiden samanaikaisten kohtaamisten määrä, eli kuinka monessa organisaatiossa henkilö X on ollut samaan aikaan kuin joku muu kohderyhmän toimija. Kun vertaillaan näitä samanaikaisuuksia, saadaan jonkinlaista kuvaa siitä, mitkä henkilöt ovat olleet uransa aikana sellaisissa kontaktirajapinnoissa, jotka ovat voineet olla merkittäviä.

On kuitenkin selvää, että käsillä olevan aineiston perusteella ei voida tehdä kovin pitkälle meneviä johtopäätöksiä. Aineiston heterogeenisyyden ja väljyyden vuoksi tulkinnot jäävät hyvin arvelevalle tasolle. Tarkoituksena on ennemminkin esitellä menetelmän käyttöä kuin esitellä analyysin tuloksia.

Toimijaverkoston mallintaminen aloitetaan määrittelemällä kaksi graafia. Ensiksi määritellään koordinaattimatriisi G_{xy} , joka kertoo kunkin solmun sijainnin kaksiuotteisessa koordinaatistossa. Koordinaatit määritetään, jotta graafien visualisoinneista saadaan yksikäsitteiset. Mielivaltainen graafi ei itsessään tarvitse koordinaatti-



Kuva 3.11: Nelisolmuinen graafi - yhteydet

informaatiota, koska vain solmujen välisillä yhteyksillä on merkitystä.

Graafin solmut ovat tässä tapauksessa kohderyhmän henkilöt. Koordinaatit voisivat olla satunnaisia, jolloin myös graafin muoto olisi satunnainen. Tärkeintä tässä analyysivaiheessa on selvittää toimijoiden välisiä mahdollisia yhteyksiä, ei niinkään muodostuvan graafin topologista muotoa. Kuitenkin jotta graafin visualisoinnin informaatioarvo olisi suurempi, on järkevää kiinnittää koordinaatiston akselit jotenkin. Päätettiin käyttää akateemista kokemusmittaria x-akselina ja yrityskokemusta y-akselina. Näin saadaan lisäinformaatiota graafien kuvaajiin.

Lisäksi tarvitaan ns. naapuruusmatriisi G_n , jonka koko on $N \times N$. Se määrittää jokaiselle graafin solmulle i , onko siitä yhteys toiseen solmuun j . Jos solmusta i on yhteys j :n, matriisin alkio g_{ij} saa arvon 1, muutoin $g_{ij} = 0$. Havainnollistuksen vuoksi kuvassa 3.11 on esimerkki yksinkertaisesta neljäsolmuisesta graafista. Graafin koordinaattimatriisi on esitetty kaavassa 3.7 ja naapuruusmatriisi kaavassa 3.8.

$$G_{xy} = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ x_3 & y_3 \\ x_4 & y_4 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 1 & 2 \\ 2 & 2 \end{pmatrix} \quad (3.7)$$

$$\mathbf{G}_n = \begin{pmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{21} & g_{22} & g_{23} & g_{24} \\ g_{31} & g_{32} & g_{33} & g_{34} \\ g_{41} & g_{42} & g_{43} & g_{44} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \quad (3.8)$$

Esimerkkimatriisissa on siis neljä ($m = 4$) symmetrisesti sijoitettua solmua, joiden sijainti on määritetty koordinaatistomatriisissa. Kaavassa 3.8 olevasta naapuruusmatriisista nähdään yhteydet eri solmujen välillä. Esimerkiksi voidaan nähdä, että solmujen 1 ja 2 välillä on yhteys, koska alkio g_{12} on ykkönen. Naapuruusmatriisi on symmetrinen, koska pätee $g_{ij} = g_{ji}$. Käytännössä sovelluksessa tämä tarkoittaa että yhteyksien suunnilla ei ole merkitystä, tai että yhteydet ovat kaksisuuntaisia. Jos henkilöt A ja B ovat yhtäaikaaisesti jossakin organisaatiossa, kyseessä on selkeästi symmetrinen relaatio. Diagonaalilla olevat alkiot tarkoittavat solmujen yhteyksiä itseensä. Jos kaikilla solmuilla on yhteys itseensä, eli muodostuu silmukka, graafi on refleksiivinen. Tässä analyysissä käsitellään vain refleksiivisiä graafeja laskentateknisistä syistä.

Koordinaatit on siis asetettu henkilögraafiin seuraavasti:

$$\mathbf{G}_{xy}(i, 1) = aka(i) \quad (3.9)$$

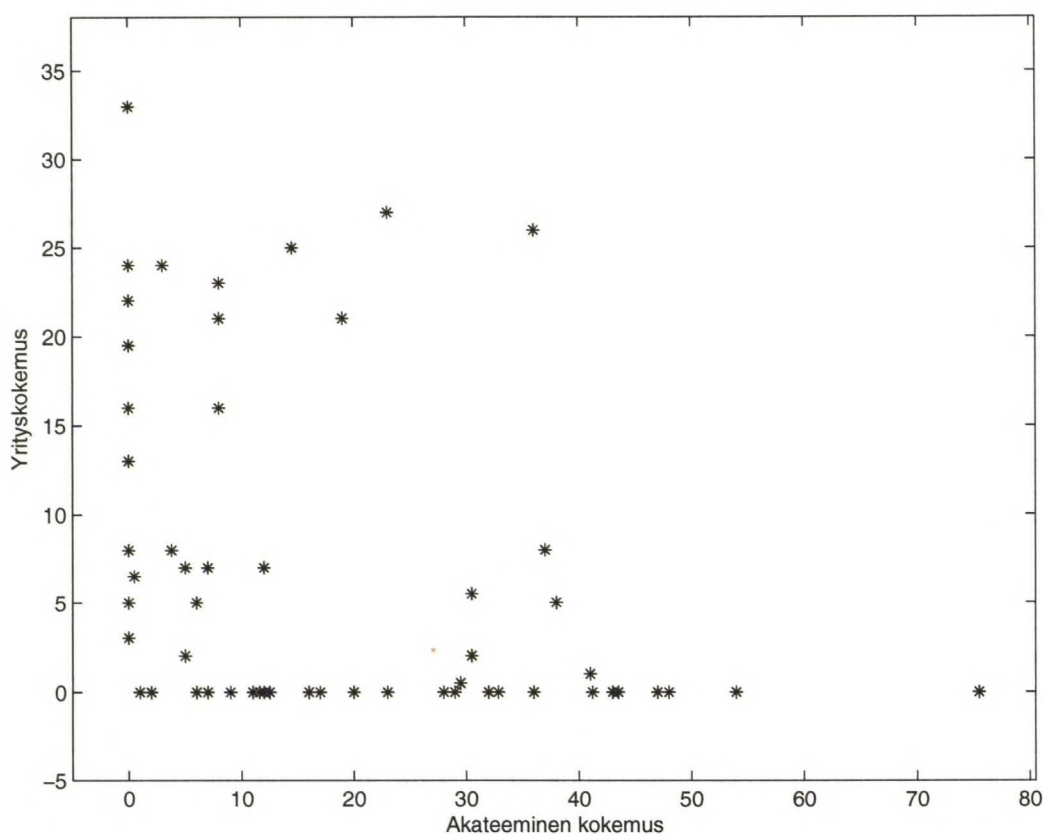
$$\mathbf{G}_{xy}(i, 2) = yrit(i) \quad (3.10)$$

Tässä $aka(i)$ tarkoittaa henkilön i akateemista kokemusta, ja vastaavasti $yrit(i)$ on yrityskokemus. Kun tämän jälkeen naapuruusmatriisin diagonaali-alkiot asetetaan ykköseksi (muut nolliksi), voidaan piirtää graafin solmut ja nähdään henkilöiden jakautuminen valitun "kokemusvaruuden" suhteen. Tämä on esitetty kuvassa 3.12.

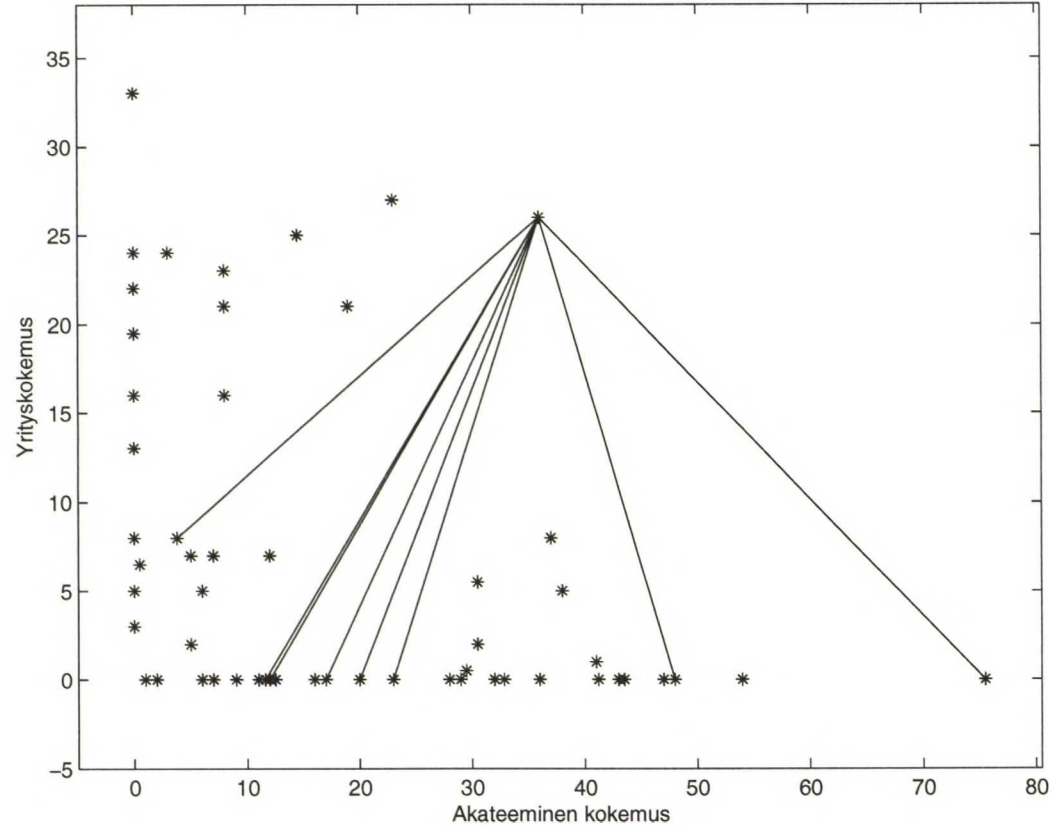
X -akselille ovat sijoittuneet puhtaasti yritysmaailman toimijat (ei akateemista kokemusta) ja vastaavasti y -akselilla ovat puhtaasti akateemiset toimijat. Akselien mittana on periaatteessa kokemus vuosissa, mutta tässä on otettava huomioon rinnakkaiset toimet, jolloin kokemusmittari voi kasvaa suuremmaksi kuin henkilön todellinen ikä on. Henkilön työura voi siis olla 20 vuoden pituinen, mutta kokemusmittari saa suurempia arvoja - esimerkiksi 35v.

Seuraavaksi verrataan jokaisen henkilön uradataa toisiin henkilöihin ja etsitään kaikki sellaiset tapaukset, joissa kaksi henkilöä ovat olleet samassa organisaatiossa samanaikaisesti. Kuvassa 3.13 olevassa graafin visualisoinnissa nähdään solmu, eli tietty henkilö X , josta on piirretty useita yhteyksiä muihin solmuihin. Nämä muut solmut ilmentävät henkilöitä, jotka ovat olleet samaan aikaan henkilön X kanssa samassa organisaatiossa. Vastaavat yhteysfrekvenssit voidaan laskea kaikille henkilöille.

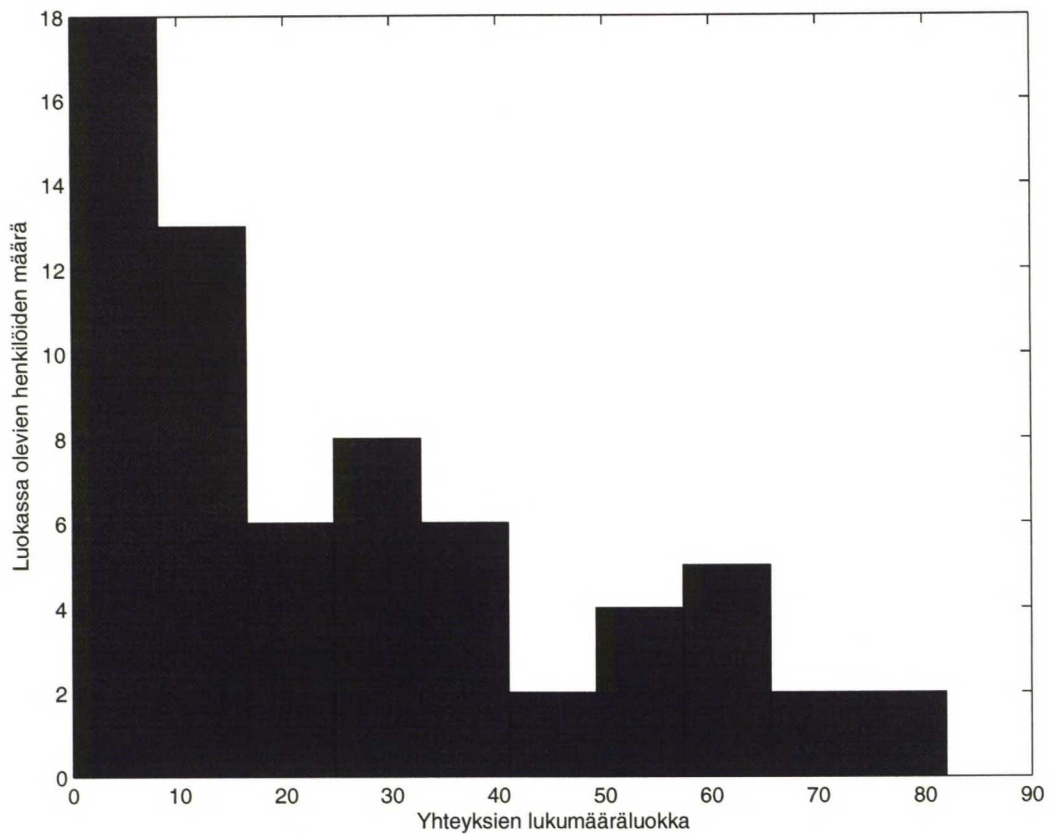
Yhteysfrekvenssien keskiarvo oli $\bar{f}_y \approx 26$ ja keskihajonta $\sigma_y \approx 22$. Henkilöillä on siis keskimäärin 26 samanaikaista organisaatioyhteyttä muiden tämän aineiston toimijoiden kanssa. Keskihajonnan melko suuri arvo ilmentää sitä, että yhteysfrekvenssit ovat keskittyneet keskiarvon ympärille. Jakauma nähdään hieman paremmin kuvasta 3.14, jossa on laskettu histogrammi yhteysfrekvensseistä.



Kuva 3.12: Henkilöt graafin solmuina



Kuva 3.13: Henkilön X yhtäaikaiset organisaatioyhteydet



Kuva 3.14: Histogrammi henkilöyhteyksistä

Kiinnostavia tässä yhteydessä ovat toimijat, jotka sijoittuvat histogrammikuvaajas-
sa äärioikealle. Nämä henkilöt voitaisiin identifioida aineistosta ja tutkia korreloi-
ko yhteyksien suuri määrä muiden tekijöiden kanssa. Tosin jokainen henkilö tulisi
tarkastella erikseen ja mahdollisesti vakioda ikä- ja muut vinouttavat tekijät tulok-
sista. Henkilöitä, joilla oli yhteyksien lukumäärä suurempi kuin 50, oli 18% koko
aineistosta. Toisaalta kiinnostavan vertailukohdan toisivat ne, joilla yhteyksiä oli
alle 10. Näitä oli 32% koko aineistosta.

Tämän jälkeen käytiin läpi yhteydet toimijoiden tyyppien mukaan. Pelkästään yri-
tyskokemusta omaavien henkilöiden yhteyksien määrä suhteessa kokonaisuuteen
oli $\sim 6\%$. Pelkästään akateemista kokemusta omaavilla yhteyksien määrä oli \sim
62% koko aineistosta. Tämän tuloksen selittää yritysten pieni määrä ja koko suh-
teessa yliopistoihin. Lisäksi on oletettavaa, että liikkuvuus uuden teknologian alan

yrityksissä on suurta, kun taas yliopistovirat ovat pidempiä.

Seuraavaksi otettiin sama verkosto tarkasteluun toisesta näkökulmasta. Selvitettiin, mitkä ovat sellaisia organisaatioita, joissa urapolkujen yhtymäkohtia on merkittävästi. Tässä tarkastelussa havaittiin, että suurin osa yhteyksistä, eli 98% oli yliopistojen tai julkisten tutkimuslaitosten sisällä. Loput, eli 2% oli yritysten sisällä tapahtuneita samanaikaisuuksia.

Luku 4

Tutkimuksen arviointi

4.1 Tavoitteiden arviointi

Ensimmäisenä tutkimustavoitteena oli tutkia, miten ansioluettelopohjaista dataa voidaan käyttää urapolkujen tutkimisessa. Tätä tehtävää varten luotiin analyysisovellus, joka koostui pääasiassa kahdesta osasta: urapolkujen tyyppejä kartoittavasta pääkomponenttianalyysistä sekä verkostanalyysistä.

PCA-analyysi oli hyvä keino redusoida aineistoa ja selvittää olennaiset piirteet urapoluista. Muuttujien valintaan on kiinnitettävä erityistä huomiota, jotta alkutilanteen kvantifioidusta, monitahoisesta urapolkuaineistosta saadaan rakennettua havaintomatriisi, joka sisältää riittävän informaation. Riippuen tutkimuskysymyksestä olisi hyvä kokeilla erilaisia muuttujakombinaatioita havaintomatriisin muodostamisessa. Käytetty klusterointitapa, joka perustui euklidisten etäisyyksien mittaamiseen ja hierarkkisen binääripuun konstruointiin, oli tehokas ja suoraviivainen tapa hahmottaa PCA-analyysin tuloksien rakennetta.

Verkostanalyysi osoittautui hyväksi tavaksi visualisoida ja kartoittaa toimijoiden välisiä yhteyksiä ja yhtymäkohtia urapoluilla. Analyysitapa asettaa kuitenkin suuria haasteita aineistolle, koska muuttujien arvojen tulee olla riittävän yhteismittalaisia, jotta tuloksista tehtävät johtopäätökset olisivat mielekkäitä. Erityisesti tämän aineiston puitteissa yhteyksien kartoittaminen ei anna kovin käyttökelpoista infor-

maatiota, koska organisaatioiden luokat vaihtelevat niin paljon. Tämä johtuu siitä, ettei riittävän tarkkoja ura- ja organisaatiotietoja ollut mahdollista saada kattavasti.

Yhteydet eivät myöskään ilmennä konkreettista yhteistyötä. Ne osoittavat toimijoiden samanaikaisuutta eri organisaatioissa, ja tämä analyysi on täten enemmänkin pohjana mahdollisille jatkotutkimusten tarkasteluille. Lisäksi on syytä huomauttaa, että kyseessä ovat formaalit urapolut, eivätkä tässä näy sellaiset verkostomaiset projektit, joita ei urapolkuaineistossa ole eksplisiittisesti määritelty.

Verkostoanalyysiä voitaisiin hyödyntää paremmin keräämällä tarkempaa henkilöiden tai organisaatioiden suhteita ilmentävää tietoa. Käsillä olevan aineiston perusteella toimijoiden väliset suhteet jäivät melko ylimalkaisiksi.

Matlab -ohjelmisto osoittautui erittäin hyväksi ympäristöksi analyysisovelluksen kehittämiseksi. Sen laajat ohjelmointimahdollisuudet tarjoavat monipuoliset puitteet erilaisten analyysien tekemiselle ja tässä tutkimuksessa kuvattujen analyysisovellusten jatkokehitykselle.

Aineiston hankinta ei onnistunut suunnitelmien mukaisesti, vaan jouduttiin tyytymään huomattavan pieneen ja heterogeeniseen näytteeseen. Aineiston heikko saatavuus ei mahdollistanut otoksen poimimista satunnaisotannalla, vaan näyte poimittiin saatavilla olevasta materiaalista. Aineisto osoittautui kuitenkin toimivaksi menetelmien testaamisessa. On syytä huomata, että tässä analyysissä käytettävät muuttujatyypit ja toisaalta verkostoanalyysin muuttujat ovat alustavia esimerkkejä. Jos halutaan vaikkapa testata jotakin hypoteesia, on muuttujat valittava sen mukaan, mitä aineistosta ollaan etsimässä.

Kvantifiointisovelluksen kehittäminen vähensi mekaanista ja virhealtista työtä aineiston tallennuksessa, ja samalla luotiin systemaattinen toimintatapa aineiston käsittelyyn ja hallintaan. Aineiston tallennus kootusti tietokantaan on hyvä menetelytapa mahdollista sovelluksen jatkokehitystä silmälläpitäen. Sovellusta voidaan kehittää siten, että eri työpisteissä työskentelevät tutkijat voivat syöttää aineistoa yhteiseen tietokantaan helposti.

Käsillä oleva aineisto ei ole riittävän kattava eikä tilastollisesti luotettava, jotta siitä voitaisiin tehdä johtopäätöksiä laajempaan ryhmään ja tämän tutkimuksen tarkoi-

tuksena olikin kehittää kvantitatiivista urapolkuanalyysiä, ei niinkään tutkia jotakin tiettyä hypoteesia. Tässä onnistuttiin hyvin.

4.2 Jatkotutkimusmahdollisuudet

Tässä tutkimuksessa kehitetty analyysikehikko avaa paljon mahdollisuuksia jatkotutkimuksen tekoon. Kun menetelmäkehikko, luokittelu- sekä analyysiohjelmisto ovat olemassa, voidaan paremmin keskittyä hypoteesien testaamiseen. Lisäksi voisi olla mielekästä kerätä haastatteluilla tukea kvantitatiivisen analyysin rinnalle.

Tämän tutkimuksen selkeänä puutteena on aineiston heterogeenisyys ja vinoutumat. Suuremmasta homogeenisesta joukosta tehtynä tutkimus olisi luotettavampi. Lähtötilanteessa perusjoukon koon on oltava suuri, mutta kuitenkin selkeästi määriteltävä. Tämän pohjalta voitaisiin tehdä otoskehys, josta saadaan satunnaisotannalla perusjoukkoon yleistettävissä oleva otos. Jos kato pysyy riittävän pienenä, saadaan tuloksista mielekkäitä tilastollisia yleistyksiä. Ongelmana tässä on kuitenkin Suomen mittakaavassa merkittävien toimijoiden vähäinen määrä, johon liittyy myöskin varsin suuri aineistokadon riski.

Luonnollinen jatkotutkimuskohde olisi suomalainen bioteknologia-ala paremmalla otoksella niin, että voitaisiin tehdä luotettavia tilastollisia yleistyksiä koko perusjoukkoon. Suomalainen bioteknologiakenttä on vielä varsin vähän tutkittua aluetta, ja laajemmalla selvityksellä alan asiantuntijoiden uratyypeistä sekä organisaatio- liikkuvuudesta voitaisiin saada arvokasta tietoa.

Vertailevat tutkimukset voisivat auttaa hahmottamaan esimerkiksi tiede- ja teknologiakulttuurin eroista johtuvia vaikutuksia bioteknologia-alan kehitykseen eri maissa. Eräs mahdollisuus voisi olla vertaileva tutkimus Suomen ja jonkun muun maan bioteknologia-alan eroista. Toinen mahdollinen jatkotutkimuksen tema voisi olla kansainvälistyminen eräänä menestystekijänä bioteknologia-alalla. Kolmas esimerkki voisi olla yliopistomaailma, jossa voitaisiin tutkia esim. tutkijoiden julkaisujen määrän korrelointia verkostojen tiheyteen ja menestykseen.

Luku 5

Yhteenveto

Tässä diplomityössä laadittiin urapolkujen analyysimenetelmä, joka pohjautui pääkomponenttianalyysin ja verkostanalyysin soveltamiseen, ja toteutettiin Matlab-ohjelmiston avulla. Työ oli luonteeltaan kartoittavaa, ja pääpainoalueeksi muodostui menetelmien kehittäminen.

Työn ensimmäisessä luvussa käsiteltiin biotekniikan ja bioteknologian määritelmiä. Bioteknologia nähtiin tässä tutkimuksessa laajempänä teknologisena verkostona ja biotekniikka suppeampana käsitteenä, joka viittaa yksittäisiin tekniikoihin ja tieteenalojen sovelluksiin. Lisäksi ensimmäisessä luvussa kuvattiin bioteknologia-alaa Suomessa sekä selvitettiin uratutkimusten kenttää.

Työn toisessa luvussa selvitettiin tutkimuksen keskeiset käsitteet, sekä kuinka urapolku määritellään tämän työn piirissä. Sitten todettiin, että Suomessa ei olla juurikaan tehty tämän alueen tutkimusta, eli on tarve tämän tyyppisille tutkimuksille. Seuraavaksi määriteltiin tutkimustehtävä kolmeosaisena:

1. Miten kvantitatiivista ansioluettelopohjaista urapolkuaineistoa voidaan käyttää bioteknologia-alan urapolkujen tutkimuksessa?
2. Hankitaan aineisto ja arvioidaan sen käytettävyyttä tutkimuksessa sekä kvantifioidaan se sopivilla työkaluilla.
3. Suoritetaan kvantitatiivinen analyysi ja tehdään seuraavia havaintoja tulok-

sista: mikä on tyypillinen aineistossa esiintyvä urapolku, ja minkälaisia verkostoja aineiston pohjalta voidaan muodostaa?

Työn kolmannessa osassa suoritettiin varsinainen analyysi lähtien aineiston hankinnasta. Aineisto hankittiin kartoittamalla suomalaisia merkittäviä bioteknologia-alan toimijoita, ja pyydettiin heiltä ansioluettelo tai vastaavat tiedot. Aineistoa saatiin yhteensä 66 toimijan osalta, joista 26 oli yrityspuolelta ja 40 julkiselta puolelta, eli yliopistoista ja julkisista tutkimuslaitoksista.

Tutkimuksen ja aineiston luotettavuutta arvioitiin systemaattisesti, koska tämän voidaan katsoa kuuluvan kehitetyn analyysiprosessiin olennaisesti, ja niin ikään antaa mallin mahdollisten jatkotutkimusten näkökulmille. Hypoteesien testauksessa aineiston kattavuuteen ja tarkkuuteen olisi kiinnitettävä erityistä huomiota.

Aineisto esikäsiteltiin numeeriseen muotoon, jotta sitä voitiin käyttää analyysiohjelmiston syötteenä, joten tässä yhteydessä esitettiin tarvittavien muuttujien valinta ja perustelut. Lisäksi kvantifiointia varten luotiin luokittelusovellus, joka helpottaa olennaisesti muuten täysin käsityönä tapahtuvaa arvojen kirjaamista tietokoneelle.

Seuraavaksi esitettiin perustulokset kohdejoukosta. Koulutustason osalta valtaosalla henkilöistä oli tohtorin tutkinto. Suurimmalla osalla oli koulutusalan jokin kemian ala, mutta myös farmasia ja lääketieteet olivat merkittävästi edustettuina.

Pääkomponenttianalyysi oli tutkimuksen toinen päämenetelmä, jossa pyrittiin hahmottamaan urapolkuaineistosta keskeisiä uratyyppejä. Jokaisesta urapolusta oli muodostettu viisi komposiittimuuttujaa, joita olivat ulkomaankokemus, akateeminen kokemus, yrityskokemus, tutkimuskokemus ja johtamiskokemus. Jokainen henkilö sai tietyn arvon kaikille näille muuttujille. Henkilöt voitiin siis asettaa viisiulotteiseen avaruuteen, mutta koska tämän suoraviivainen visualisointi on erittäin vaikeaa, käytettiin pääkomponenttianalyysiä redusoimaan aineisto viisiulotteisesta avaruudesta kaksiulotteiseen. Tämän jälkeen tulokset voitiin esittää kaksiulotteisessa koordinaatistossa. Analyysin tuloksista voitiin löytää kolme klusteria, joista suurin sai nimen "Tyypillinen", kaksi pienempää saivat nimikseen "Akateeminen" ja "Yrityskenttä". Analyysimenetelmällä saatiin siis etsittyä tyypillisin urapolkuluokka aineistosta.

Verkostoanalyysissä etsittiin henkilöiden ammatillisia kohtaamispaikkoja urapoluilla. Lähtökohtana luotiin graafi, jonka solmuina olivat henkilöt sijoitettuna koordinaatistoon, jonka toisena akselina oli akateeminen kokemus ja toisena akselina yrityskokemus. Tämän jälkeen laskettiin henkilöiden frekvenssejä siitä, kuinka monen henkilön kanssa he olivat olleet samassa organisaatiossa samaan aikaan. Tämän ajateltiin ilmentävän mahdollista yhteyttä henkilöiden välillä, mutta aineiston rajoituksen vuoksi tuloksia on ajateltava vain menetelmän testituloksina. Suurin ongelma verkostoanalyysissä oli organisaatioiden varsin väljä luokitus johtuen lähtöaineiston väljyydestä ja toisaalta luokittelun ongelmista.

Työn arviointiosuudessa, luvussa neljä, käytiin läpi tutkimuksen tavoitteet ja todettiin, että työ on saavuttanut melko hyvin tavoitteet aineistovaikeuksista huolimatta. PCA-analyysi osoittautui hyväksi menetelmäksi redusoidessa moniulotteista urapolkuaineistoa ja etsittäessä klustereita. Havaintomatriisin muodostamiseen on kiinnitettävä huomiota, jos testataan jotakin hypoteesia aineiston pohjalta. Verkostoanalyysi nähtiin hyvänä yhteyksien visualisoinnissa ja etsittäessä yhteysfrekvenssejä, mutta se vaatisi tarkempaa aineistoa toimijoiden tai organisaatioiden välisistä yhteyksistä. Matlab-ohjelmisto osoittautui hyväksi välineeksi varsinaisen kvantitatiivisen analyysiprosessin kehittämisessä ja laskennassa. Lopuksi esitettiin muutamia jatkotutkimusmahdollisuuksia, joista esimerkkinä vertailu Suomen ja jonkin muun maan bioteknologiakentän eroista.

Lähteet

- Leo Ahlstedt. *Erikoistuminen ja liikkuvuus liikkeenjohtajan urakehitystekijöinä*. Helsingin Kauppakorkeakoulun julkaisuja, 1978.
- Tapani Alkula, Seppo Pöntinen ja Pekka Ylöstalo. *Sosiaalitutkimuksen kvantitatiiviset menetelmät*. Werner Söderström Osakeyhtiö, 1994.
- Alan T. Bull, Geoffrey Holt ja Malcolm D. Lilly. Biotechnology - international trends and perspectives. 1982. <http://www.oecd.org/dataoecd/34/9/2097562.pdf>, haettu 15.4.2004.
- Michael Callon. Society in the making: The study of technology as a tool for sociological analysis. Teoksessa *The Social Construction of Technological Systems*, sivut 83–103. The MIT Press, 1999.
- Manuel Castells. *The rise of the network society*. Oxford: Blackwell Publishers, 1996.
- Arlene Fink. *The Survey Handbook*. Sage Publications, 1995.
- Saara Hassinen. Life sciences and biotechnology at the crossroads. *Kemia-Kemi Vol. 29*, sivut 19–20, 2002.
- M. Höyssä, H. Bruun ja J. Hukkinen. The co-evolution of social and physical infrastructure for biotechnology innovation in turku, finland. *Research Policy, Vol. 33*, 2004.
- Jan-Erik Johanson, Mikko Mattila ja Petri Uusikylä. *Johdatus verkostanalyysiin*. Kuluttajatutkimuskeskus, 1995.

- Terttu Luukkonen toim. *Bioteekniikka - tietoon perustuvaa liiketoimintaa*. Elinkeinoelämän tutkimuslaitos, ETLA, 2004.
- Riikka Mantila. Turun pharmacy vihittiin käyttöön. *Yliopistotiedot. Turun yliopisto*. <http://www.utu.fi/yt/7-2002/18.html>, 2002. Julkaistu myös Turun yliopisto. *Yliopistotiedot*. 41. vuosikerta 27. toukokuuta 2002, numero 7/2002.
- Susan Milton ja Jesse Arnold. *Introduction to Probability and Statistics*. McGraw-Hill, Inc., 1995.
- John Scott. *Social Network Analysis, A Handbook*. Sage Publications, 2000.
- Subhash Sharma. *Applied Multivariate Techniques*. John Wiley & Sons, Inc, 1996.
- Juhani Taskinen. *Arvio biotekniikan tutkimuksen hyödyntämisestä Suomessa*. pro gradu, Jyväskylän yliopisto, 2003.
- Stanley Wasserman ja Katherine Faust. *Social Network Analysis*. Cambridge University Press, 1999.
- Mark Allen Weiss. *Data Structures and Algorithm Analysis in C*. Addison Wesley, 1997.

Liite A

Matlab- analyysiprosessien lähdekoodi

Tässä liitteessä on esitetty analyysiprosesseissa käytetty Matlab-kielinen lähdekoodi. Työssä käytettiin Matlab-ohjelmiston versiota 6.1.0.450 (R12.1). Käyttöjärjestelmänä oli Debian GNU/Linux.

```
% -----  
% Analyysin päätaso (run.m)  
% -----  
  
% Ladataan data ja esitetään perustiedot kohderyhmästä  
initialize;  
compute_basic;  
show_basic;  
  
% Lasketaan kokemusmittarit ja näytetään graafit  
% Tehdään pca-analyysi ja klusterointi  
compute_experiences;  
show_experiences;  
  
pca;  
cluster_pca;  
  
% Lasketaan verkostanalyysi ja esitetään tulokset  
compute_networks;  
show_networks;
```

```
% -----
% Tietojen luku ja alustukset (initialize.m)
% -----

disp('Ladataan uradata, määrittelyt ja lasketaan perustiedot');
clear;

% Luetaan datatiedosto sisään, lasketaan koko
data_66          % Sisältää varsinaisen datan
class_names      % Sisältää luokkien nimet

N = length(data);
Y = 2003;

% Asetetaan joitakin vakioita luokitus
ACADEMIC=1;
FIRM=3;
LEAD=5;
RESEARCH=3;

profession_names = {
    'N/A' ...
    'Biokemia' ...
    'Farmasia yms.' ...
    'Lääketieteet' ...
    'Molekyyli/mikrobiologia' ...
    'Muu bioala' ...
    'Muu kemian ala' ...
    'Muut' ...
};

org_names = {
    % 'N/A' ...
    'Yliopisto' ...
    'Muu julkinen organisaatio' ...
    'Yritys' ...
    'Yhdistys/järjestö'
};

posit_names = {
    'N/A' ...
    'Johto' ...
    'Keskijohto' ...
    'Asiantuntija' ...
    'Toimihenkilö' ...
    'Työntekijä' ...
    'Muu'
};
```



```
% Urapolkujen pituudet career_lens -muuttujaan

for k=1:N,
    career_lens(k) = length(data(k).durations);
end

disp('Valmis.');
```

```
% -----  
% Perustietojen laskenta (compute_basic.m)  
% -----  
  
disp('Lasketaan kohdejoukosta perustunnuslukuja...');  
  
% Ikätietoja  
ages = (ones(N,1)*Y)' - [data.birth_year];  
  
agehist = histc(ages, [30 35 40 45 50 55 60 65 ] );  
agehist = agehist(1:7);  
  
% Esitetään jakauma koulutuksesta  
profhist = histc([ data.profession ], [0 1 2 3 4 5 6 7] );  
  
% Missä organisaatiossa henkilöt töissä tutkimushetkellä?  
for k=1:N,  
    [dummy last_index] = size(data(k).organization_types);  
    last_org(k) = data(k).organization_types(last_index);  
end  
  
lasthist = histc(last_org, [1 2 3 4] );  
  
% Missä tehtävässä henkilöt ovat viimeisessä työpaikassa?  
for k=1:N,  
    [dummy last_index] = size(data(k).positions);  
    last_positions(k) = data(k).positions(last_index);  
end  
  
posithist = histc(last_org, [1 2 3 4] );  
  
for k=1:4,  
    positionhist(k) = round(100 * posithist(k)/N);  
end  
  
disp('Valmis.');
```

```
% -----  
% Kokemusmuuttujien laskenta (compute_experiences.m)  
% -----  
  
disp('Lasketaan kokemusmuuttujat...');  
  
% Lasketaan ulkomaankokemusmittari (vuosissa)  
for k=1:N,  
    exp_abroad(k) = 0;  
  
    for i=1:career_lens(k),  
        if data(k).nationalities(i) == 2  
            exp_abroad(k) = exp_abroad(k) + data(k).durations(i);  
        end  
    end  
end  
  
% Akateeminen työkokemus (vuosissa)  
for k=1:N,  
    exp_academia(k) = 0;  
  
    for i=1:career_lens(k),  
        if data(k).organization_types(i) == ACADEMIC  
            exp_academia(k) = exp_academia(k) + data(k).durations(i);  
        end  
    end  
end  
  
% Kokemus yritysmaailmassa (vuosissa)  
for k=1:N,  
    exp_firm(k) = 0;  
  
    for i=1:career_lens(k),  
        if data(k).organization_types(i) == FIRM  
            exp_firm(k) = exp_firm(k) + data(k).durations(i);  
        end  
    end  
end  
  
% Johtotehtävät (vuosissa)  
for k=1:N,  
    exp_lead(k) = 0;  
  
    for i=1:career_lens(k),  
        if data(k).post_duties(i) == LEAD  
            exp_lead(k) = exp_lead(k) + data(k).durations(i);  
        end  
    end  
end
```

```
end

% Tutkimustehtävät
for k=1:N,
    exp_research(k) = 0;

    for i=1:career_lens(k),
        if data(k).post_duties(i) == RESEARCH
            exp_research(k) = exp_research(k) + data(k).durations(i);
        end
    end
end

disp('Valmis.');
```



```
% -----  
% Pääkomponenttianalyysin laskenta (pca.m)  
% -----  
  
disp('Lasketaan PCA-analyysi muutamien muuttujien suhteen');  
  
% Koulutusalat  
professions = [ data.profession ];  
mobility = [ career_lens ];  
  
for k=1:N,  
    org = unique(data(k).organizations);  
    [dummy num] = size(org);  
    num_organizations(k) = num;  
end  
  
% Muodostetaan havaintomatriisi PCA-analyysia varten  
A = [ transpose(exp_abroad) ...  
      transpose(exp_academia) ...  
      transpose(exp_firm) ...  
      transpose(exp_research) ...  
      transpose(exp_lead) ...  
  
      ];  
  
Astd = std(A);  
sr = A ./ repmat(Astd, 66, 1);  
[pcs,newdata,variances,t2] = princomp(sr);  
  
plot(newdata(:,1),newdata(:,2),'bo');  
grid on;  
xlabel('1. Principal Component');  
ylabel('2. Principal Component');  
  
disp(''  
pcs(:,1:3)  
  
disp('Kuvaaja PCA-analyysin tuloksesta.');
```

```

% -----
% PCA tulosten klusterointi (cluster_pca.m)
% -----

disp('Klusteroidaan PCA-analyysillä saatu 2D-data:');

X = [newdata(:,1) newdata(:,2)];
Y = pdist(X);
Z = linkage(Y, 'average');
T = cluster(Z,4);

disp('Dendrogrammi datasta:');
dendrogram(Z, 66);
ylabel('Klusterointitaso');
xlabel('Henkilöt');
set(gca,'XTickLabel',' ');
pause;

hold off;
disp('Klusterien koot seuraavassa:');
histc(T, [0 1 2 3 4])
disp('Piiirretään neljä klusteria, kukin eri värillä...');

colors = char('bo', 'ro', 'ko', 'mo');

c1 = find(T==1);
c2 = find(T==2);
c3 = find(T==3);
c4 = find(T==4);

for i=1:4,
    plot(newdata(find(T==i), 1), newdata(find(T==i), 2), colors(i,:));
    hold on;
end
grid on;
hold off;

disp('Lasketaan muutamia keskiarvoja klusterien pohjalta');
disp('exp_abroad:');
com1 = [sum(exp_abroad(c1))/length(c1) ...
        sum(exp_abroad(c2))/length(c2) ...
        sum(exp_abroad(c3))/length(c3) ...
        sum(exp_abroad(c4))/length(c4)
        ]

disp('exp_academia:');
com2 = [sum(exp_academia(c1))/length(c1) ...
        sum(exp_academia(c2))/length(c2) ...

```

```
        sum(exp_academia(c3))/length(c3) ...
        sum(exp_academia(c4))/length(c4)
    ]

disp('exp_firm:');
com2 = [sum(exp_firm(c1))/length(c1) ...
        sum(exp_firm(c2))/length(c2) ...
        sum(exp_firm(c3))/length(c3) ...
        sum(exp_firm(c4))/length(c4)
    ]

disp('exp_research:');
com2 = [sum(exp_research(c1))/length(c1) ...
        sum(exp_research(c2))/length(c2) ...
        sum(exp_research(c3))/length(c3) ...
        sum(exp_research(c4))/length(c4)
    ]

disp('exp_lead:');
com2 = [sum(exp_lead(c1))/length(c1) ...
        sum(exp_lead(c2))/length(c2) ...
        sum(exp_lead(c3))/length(c3) ...
        sum(exp_lead(c4))/length(c4)
    ]

disp('Valmis.');
```

pause;

```

% -----
% Verkostanalyysin laskenta (compute_networks.m)
% -----

disp('Lasketaan verkostanalyysiä, tämä vie jonkin aikaa...');

% person_links: Henkilöiden kohtaamispaikat. Kun A:lla ja B:llä
% samaan aikaan työpositio samassa paikassa, lisätään A:n laskuria.

% organization_links: organisaatiofrekvenssit. Kun A ja B samaan
% aikaan samassa organisaatiossa X, lisätään org_connection_frew

person_links = zeros(1, N);          % hlöiden koht. pisteet 1-hlön
organization_links = zeros(1, ...    % organisaatiofrekvenssit 1-orgN
max([data.organizations]));

% Määritetään koordinaatit yritys- ja tutkimuskokemuksen mukaan
for i=1:N,
    G_xy_matrix(i,1) = exp_academia(i);
    G_xy_matrix(i,2) = exp_firm(i);
end

% Asetetaan ykköset diagonaalilinjalle (NxN-matriisi)
G = diag(ones(1,N));

if 0, ...
for i=1:N,
    for j=1:N,
        if (i == j)
            G(i, j) = 1;
        else
            G(i, j) = 0;
        end
    end
end
end

G_full = G;
G_person1 = G;
G_org = G;

for i=1:N,          % henkilöt läpi
    for j=1:N,      % vertaillen rinnan
        if i == j
            continue; % ei verrata samaa hlöä
        end

        [dummy org_num_i] = size(data(i).organizations);

```



```

[dummy org_num_j] = size(data(j).organizations);

for k=1:org_num_i,          % henkilön i organisaatiot
    for l=1:org_num_j,      % henkilön j organisaatiot

        if data(i).organizations(k) == data(j).organizations(l)
            i_first = data(i).begin_years(k);
            i_last = data(i).durations(k) + i_first;

            j_first = data(j).begin_years(l);
            j_last = data(j).durations(l) + j_first;

            if ((j_first >= i_first) & (j_first <= i_last)) | ...
                ((j_last >= i_first) & (j_last <= i_last))

                person_links(i) = person_links(i) + 1;
                organization_links(data(i).organizations(k)) = ...
                    organization_links(data(i).organizations(k)) + 1;

            if (i==7)
                seven(person_links(i),:) = ...
                    [j data(i).organizations(k) data(j).organizations(l) ...
                     i_first i_last j_first j_last];

                seven_pers(person_links(i)) = j;
            end

            G_full(i, j) = G_full(i, j) +1;
            if (i==1)
                G_person1(i,j) = G_person1(i,j)+1;
            end
            if (k==1)
                G_org(i,j) = G_org(i,j)+1;
            end
        end
    end
end

end % org_num_j
end % org_num_i
end % j

if mod(i, 3)==0,
    s = sprintf('%d %s', round(100*i/N), '%');
    disp(s);
end

end % i

disp('Valmis.');
```

```
% -----
% Kohdejoukon perustietojen esitys (show_basic.m)
% -----

disp('Kuvaillaan kohdejoukkoa, perustiedot...');

% Ikätietoja
disp('Minimi- ja maksimi- iät sekä keskiarvo');
min(ages)
max(ages)
sum(ages)/length(ages)
disp('Paina enter..'); pause;

disp('Piirakkakuvaaja ikäjakaumasta');
agehist
pie(agehist, {'30-35' '35-40' '40-45' '45-50' '50-55' '55-60' '60-65'} );
title('Kohderyhmän ikäjakauma');
disp('Paina enter..'); pause;

% Esitetään jakauma koulutuksesta
disp('Piirakkakuvaaja koulutusjakaumasta');
profhist
pie(profhist, profession_names);
title('Kohderyhmän koulutusjakauma');
disp('Paina enter..'); pause;

% Missä organisaatiossa henkilöt töissä tutkimushetk.
disp('Piirakkakuvaaja organisaatioista, joissa henkilöt töissä');
lasthist
pie(lasthist, org_names);
title('Organisaatiot');
disp('Paina enter..'); pause;

% Missä tehtävässä henkilöt ovat viimeisessä postissa
disp('Johto, keskijohto, asiantuntija, toimihenkilö (prosentit):');
positionhist

disp('Valmis.');
```

```
% -----  
% Laskettujen kokemusmittarien esittäminen (show_experiences.m)  
% -----  
  
disp('Näytetään erilaisia kokemusmittareita');  
  
disp('Näytetään ulkomaankokemus vuosissa:');  
hist(exp_abroad,20)  
title('Ulkomaankokemus vuosissa')  
ylabel('Vuodet')  
ylabel('Henkilomaara')  
xlabel('Kokemus vuosissa')  
pause;  
  
disp('Näytetään akateeminen kokemus vuosissa:');  
hist(exp_academia,20)  
title('Akateeminen kokemus vuosissa')  
ylabel('Vuodet')  
ylabel('Henkilomaara')  
xlabel('Kokemus vuosissa')  
pause;  
  
disp('Näytetään yrityskokemus vuosissa:');  
hist(exp_firm,20)  
title('Yrityskokemus vuosissa')  
ylabel('Vuodet')  
ylabel('Henkilomaara')  
xlabel('Kokemus vuosissa')  
pause;  
  
disp('Näytetään johtamiskokemus vuosissa:');  
hist(exp_lead,20)  
title('Johtamiskokemus vuosissa')  
ylabel('Vuodet')  
ylabel('Henkilomaara')  
xlabel('Kokemus vuosissa')  
pause;  
  
disp('Näytetään kokemus tutkimustehtävistä vuosissa:');  
hist(exp_research,20)  
title('Tutkimuskokemus vuosissa')  
ylabel('Vuodet')  
ylabel('Henkilomaara')  
xlabel('Kokemus vuosissa')  
pause;  
  
disp('Valmis.');
```

```

*
% -----
% Verkostanalyysin tulosten esittäminen (show_networks.m)
% -----

disp('Kuvataan verkostanalyysin tulokset...');

maxx = max(exp_academia)+5;
minx = -5;
maxy = max(exp_firm)+5;
miny = -5;

disp('Piirretään henkilögraafin solmut');
gplot(G, G_xy_matrix, '*-');

axis([minx maxx miny maxy])
xlabel('Akateeminen kokemus');
ylabel('Yrityskokemus');
pause;

if 0
disp('Piirretään kaikki yhteydet');
gplot(G_full, G_xy_matrix, '*-');
axis([minx maxx miny maxy])
%title('Henkilögraafi ')
xlabel('Academic');
ylabel('Firm');
pause;
end

disp('Piirretään ensimmäisen henkilön yhteydet');
gplot(G_person1, G_xy_matrix, '*-');
axis([minx maxx miny maxy])
%title('Henkilögraafi ')
xlabel('Akateeminen kokemus');
ylabel('Yrityskokemus');
pause

disp('Piirretään ensimmäisen organisaation yhteydet');
gplot(G_org, G_xy_matrix, '*-');
axis([minx maxx miny maxy])
%title('Henkilögraafi ')
xlabel('Akateeminen kokemus');
ylabel('Yrityskokemus');
pause;

% Huom. mukana on kaksinkertaiset yhteydet, jaetaan pois...
plinks = person_links/2;

```



```
disp('Yhtäaikaiset organisaatioyhteydet henkilöiden mukaan');
bar([1:N], sort(plinks), 'r')
title('Simultaneous organisation connections per person')
xlabel('Person number');
ylabel('Number of connections');
pause;

disp('Keskiarvo ja keskihajonta henkilöyhteyksistä:');
mean(plinks);
std(plinks);

disp('Histogrammi yhteyksistä');
hist(sort(plinks));
h = hist(sort(plinks));
%title('Jakauma (histogrammi) henkilöyhteyksien määrästä');
xlabel('Yhteyksien lukumääräluokka');
ylabel('Luokassa olevien henkilöiden määrä');
pause;

disp('Henkilöiden lkm, joiden yhteyksien määrä oli yli 50:');
length(find(plinks > 50))
disp('Henkilöiden lkm, joiden yhteyksien määrä oli alle 10:');
length(find(plinks < 10))

disp('Henkilöyhteyksien kokonaismäärä:');
sum(plinks);

er0 = find(exp_research==0); % ei akat. kokemusta, yrityskok.
ef0 = find(exp_firm==0); % pelkkää akat. kokemusta, ei lainkaan

disp('Puht. yritystoimijoiden linkkien määrä suhteessa kokonaisuuteen:');
sum(plinks(er0))/sum(plinks)

disp('Puht. akat. toimijoiden linkkien määrä suhteessa kokonaisuuteen:');
sum(plinks(ef0))/sum(plinks)

pause;

disp('Relevantit organisaatiot yhteyksien mukaan');
barh([1:length(organization_links)], organization_links)
axis([0 450 0 25]);
set(gca,'YAxisLocation', 'right')
set(gca,'YTick',1:25)
set(gca,'YTickLabel',Organisaatio(1:25,:))
grid on;

%title('Most relevant organizations');
ylabel('Organisaation numero');
```

```
xlabel('Yhteyksien määrä');
pause;

disp('w/o names!!!');
set(gca,'YTickLabel',1:25);

pause;

% merkittävimmät organisaatiot

orgnotzero = find(organization_links > 0);

tmps = size(Organisaatio);
strsize = tmps(2);

for orgi = orgnotzero,
    s = sprintf('%s %0.5g', Organisaatio(orgi, 1:strsize), organization_links(orgi));
    disp(s);
end

disp('Valmis.');
```

Liite B

Kuvaus luokittelutyökalusta

Tässä liitteessä esitellään analyysin aputyökaluna ollut luokittelusovellus, jota on käytetty aineiston saamiseen numeeriseen muotoon. Yksinkertainen luokittelusovellus kehitettiin aineiston koodauksen helpottamiseksi ja sen tutkimiseksi, miten kyseisestä työkalusta voidaan hyötyä tämäntyyppisissä projekteissa. Aineistomäärän ollessa näin pieni ei hyöty ole kovin suuri, mutta suuremmilla datamäärillä ja useiden tutkijoiden käytössä tämänkaltaisen räätälöity työkalu olisi erittäin hyödyllinen. Ohjelman tarkoituksena on:

- Helpottaa ja osin automatisoida aineiston luokittelua yksinkertaisen käyttöliittymän kautta.
- Tallentaa aineisto keskitettyyn tietokantaan.
- Tarjota helppo tapa muuntaa aineistoa formaatista toiseen.
- Antaa mahdollisuus monen käyttäjän yhtäaikaiseen aineiston editointiin.

Ohjelma käyttää MySQL -tietokantaa ja käyttöliittymä on web-pohjainen.

Kuvassa B.1 nähdään luokittelusovelluksen lisäystila, johon siirrytään automaattisesti avattaessa sovellus. Henkilön perustiedot, kuten nimi ja koulutus, syötetään tässä ruudussa.

Urapolkujen koodaus
[Lisää] [Luokkatiedot] [Luokat matlab] [Pikalista] [Tarkat tiedot] [Matlab-muoto] [Asetukset] [Poista kaikki]

Perustietojen lisäys
Nimi:
Syntymävuosi:
Koulutustaso: Muu mikä:
Koulutusala: Muu mikä:
Suoritusvuosi:
Urapolun pituus

Kuva B.1: Luokittelusovellus - perustietojen lisäys

Kun henkilön perustiedot on lisätty tietokantaan, syötetään yksitellen jokainen uraskel kuvan B.2 näköiselle lomakkeelle. Uusia luokkia voidaan syöttää tarpeen mukaan, ja ohjelma lisää ne automaattisesti käytettäviksi. Näin numeerisen tiedon käsittely on jätetty kokonaan sovelluksen tehtäväksi.

Luokat saadaan näkyviin sovelluksessa, kuten kuvassa B.3 nähdään, mutta ne saadaan milloin tahansa myös Matlabin ymmärtämässä muodossa valikon kohdasta *Luokat matlab*.

Kuvassa B.4 nähdään lista, josta päästään käsiksi jo syötettyjen henkilöiden tietoihin (kuvassa nimet poistettu). Tästä saadaan yleissilmäys dataan. Tarkemmat tiedot saadaan sovelluksen valikon kohdasta *Tarkat tiedot*. Edelleen valmiiksi luokitellussa Matlab-muodossa koko aineisto saadaan valikon kohdasta *Matlab-muoto*.

Urapolkujen koodaus

[Lisää] [Luokkatiedot] [Luokat matlab] [Pikalista] [Tarkat tiedot] [Matlab-muoto] [Asetukset] [Poista kaikki]

Urapolkutietojen lisäys

Lisää organisaatiotyyppi:

Lisää organisaatio:

Lisää asema:

Lisää toiminimitys:

Lisää toimen ala:

Lisää toimen päätehtävä:

Tallenna yllä olevat

Henkilö numero 67, Matti Meikäläinen

Toimi 1/5

Organisaatiotyyppi:

Organisaatio:

Toimen kansallisuus (1/2):

Asema (positio):

Toiminimitys:

Toimen ala:

Päätehtävä:

Alkuvuosi:

Kesto vuosissa:

Toimi 2/5

Organisaatiotyyppi:

Kuva B.2: Luokittelusovellus - urapolun syöttäminen

Urapolkujen koodaus		
[Lisää] [Luokkatiedot] [Luokat matlab] [Pikalista] [Tarkat tiedot] [Matlab-muoto] [Asetukset] [Poista kaikki]		
Koulutustaso	Koodi	Poista
Opistotasoinen tutkinto	1	X
Ammattikorkeakoulututkinto	2	X
Alempi korkeakoulututkinto	3	X
Ylempi korkeakoulututkinto	4	X
Lisensiaatti	5	X
Tohtori	6	X
N/A	7	X
Koulutusalan nimike	Koodi	Poista
Biokemia	1	X
Farmasia yms.	5	X
Lääketieteet	6	X
Molekyyli/mikrobiologia	3	X
Muu bioala	4	X
Muu kemian ala	2	X
Muut	7	X
N/A	8	X
Organisaatio	Koodi	Poista

Kuva B.3: Luokittelusovellus - luokat

Urapolkujen koodaus

[Lisää] [Luokkatiedot] [Luokat matlab] [Pikalista] [Tarkat tiedot] [Matlab-muoto] [Asetukset] [Poista kaikki]

ID	Nimi	Syntymävuosi	Urapolun pituus	Uradata	Tarkat tiedot	Poisto
1	(Nimet poistettu)	1945	7	ok	Näytä	X
2		1944	3	ok	Näytä	X
3		1949	16	ok	Näytä	X
4		1945	10	ok	Näytä	X
5		1944	10	ok	Näytä	X
6		1964	7	ok	Näytä	X
7		1949	9	ok	Näytä	X
8		1962	12	ok	Näytä	X
9		1950	6	ok	Näytä	X
10		1954	16	ok	Näytä	X
11		1958	11	ok	Näytä	X
12		1954	13	ok	Näytä	X
13		1950	7	ok	Näytä	X
14		1942	12	ok	Näytä	X
15		1954	8	ok	Näytä	X
16		1949	9	ok	Näytä	X
17		1950	12	ok	Näytä	X
18		1949	8	ok	Näytä	X
19		1955	6	ok	Näytä	X
20		1950	17	ok	Näytä	X
21		1941	9	ok	Näytä	X
22		1955	5	ok	Näytä	X
23		1964	6	ok	Näytä	X
24		1942	0	ok	Näytä	X

Kuva B.4: Luokittelusovellus - lista henkilöistä